AD_____

Award Number: DAMD17-97-1-7063

TITLE: Structural Studies of the pRB Tumor Suppressor Complexed with Human Papillomavirus E7 Proteins

PRINCIPAL INVESTIGATOR: Adrienne Clements, B.S.

CONTRACTING ORGANIZATION: The Wistar Institute
Philadelphia, Pennsylvania 19104

REPORT DATE: July 2000

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2000	Final (1 Jun 97 - 31 May 00)

4. TITLE AND SUBTITLE
Structural Studies of the pRB Tumor Suppressor Complexed with Human Papillomavirus E7 Proteins

5. FUNDING NUMBERS
DAMD17-97-1-7063

6. AUTHOR(S)
Adrienne Clements, B.S.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Wistar Institute
Philadelphia, Pennsylvania 19104

E-MAIL:
aclement@mail.med.upenn.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Since viral oncoproteins are expected to compete with and imitate interactions that pRB has with cyclin D1, understanding high affinity pRB-viral oncoprotein complexes will provide tremendous insight into the specific interactions required for the development of small compounds that can destabilize pRB-cyclin D1 complexes in cyclin D1-mediated breast cancer. Therefore, the primary goal of this project is to determine the three dimensional structure of pRB bound HPV E7 and Adenovirus 5 E1A. Additionally, since p53 functions as a tumor suppressor that is often inactivated in breast cancer, a secondary goal of this project is to determine the structure of the PCAF acetyltransferase domain with coenzyme A and a p53-derived peptide in order to gain insight into PCAF-mediated p53 activation

This study demonstrates that bacterially coexpressed pRB(376-792) and viral oncoproteins form complexes. To date, these purified complexes have resisted crystallization. However, crystallography of HPV1a E7 and NMR studies of Adenovirus 5 E1A have had preliminary success. Sedimentation equilibrium experiments of the pRB/viral oncoprotein complexes and their individual components have been utilized to characterize protein oligomerization states. Additionally, the crystal structure of PCAF bound to coenzyme A has been solved and provides insight into PCAF-mediated p53 acetylation and activation.

14. SUBJECT TERMS
Breast Cancer, Predoc Award
protein structure, X-ray crystallography, pRB-E7 complex, cyclin D1, structure-based drug design

15. NUMBER OF PAGES
29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

AC Where copyrighted material is quoted, permission has been obtained to use such material.

AC Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

AC Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Adrienne Clements 6/28/00

PI - Signature Date

Table of Contents

INTRODUCTION

Transcriptional activators and repressors are often involved in cell cycle control and are altered in breast cancer[1]. Consequently, this structural biology project focuses on the following proteins involved in cell cycle regulation: the retinoblastoma tumor suppressor protein (pRB), DNA viral oncoproteins HPV E7 and Adenovirus E1A, the p300/CBP-associated factor (PCAF) and p53. pRB is an example of a transcriptional repressor that is critically involved in the control of the G1-S phase transition of the cell cycle[2]. In cyclin D1- mediated breast cancer, overexpressed cyclin D1 binds to and inactivates pRB through phosphorylation, which promotes uncontrolled cell proliferation[3]. Since DNA viral oncoproteins and the cellular cyclin D protein share homologous regions that are essential for interaction with pRB, we hypothesize that these viral oncoproteins compete with and possibly imitate some interactions that pRB normally has with cyclin D and other cellular proteins. Therefore, a primary goal of this project is to perform structural studies of the pRB tumor suppressor complexed with viral oncoproteins HPV E7 and Adenovirus E1A in order to gain insight into pRB function and high affinity pRB-protein interaction. Although structural information is already available for the pRB small pocket domain bound to a nine amino acid HPV16 E7 peptide (amino acids 20-29)[4], the crystallized E7 peptide is incapable of inactivating pRB and binds to pRB with a twenty-fold weaker affinity compared with full-length HPV 16 E7[5]. Consequently, the constructs of HPV E7 and Adenovirus E1A utilized in this project include the additional pRB inactivating regions[6,7]. A second focus of this project is to elucidate the mechanism of human PCAF-mediated p53 activation using structural biology. p53 is a transcriptional activator that is also involved in the control of the G1-S phase transition of the cell cycle[8]. p53 functions as a tumor suppressor that is often mutated in breast cancer[9]. Risk of breast cancer recurrence and breast cancer related death is increased by at least 50% if p53 is abnormal[9]. Human PCAF mediates transcriptional activation through its ability to acetylate nucleosomal histone substrates as well as transcriptional activators such as the p53 tumor suppressor[10,11]. Specifically, PCAF acetylates lysine 320 of p53 *in vitro*, resulting in an increased affinity of p53 to DNA[11]. Correlatively, lysine 320 of p53 is acetylated *in vivo* in response to DNA damage[11]. PCAF is also inhibited by the Adenovirus E1A oncoprotein, which leads to the suppression of PCAF mediated transactivation[12]. Since PCAF is targeted by a viral oncoprotein and modulates p53 tumor suppressor activity, the second goal of this project is to determine the structure of the PCAF acetyltransferase domain with coenzyme A and a p53-derived peptide in order to gain insight into p53 activation.

BODY

pRB-viral oncoprotein studies

The cDNA of HPV16 E7 and HPV1a E7 was obtained from Dr. Robert Ricciardi and Dr. Thomas Iftner, respectively. Three constructs of HPV16E7 and of HPV1a E7 were subcloned into a pRSETA vector for protein expression with a T7 promoter-T7 polymerase expression system in the bacterial strain BL21(DE3). Constructs of the full length HPV16 E7 (amino acids 1-98) and full length HPV1a E7 (amino acids 1-93) were produced and include three highly conserved regions (CR1-CR3, Figure1A) among DNA viral oncoproteins HPVE7, Adenovirus E1A and SV40 large T antigen. Constructs containing the minimal pRB binding domains (CR2-CR3, Figure 1A, Construct 2) were also generated for HPV16 E7 (amino acids 17-98) and HPV1a E7 (amino acids 16-93). Smaller constructs that included only the pRB-inactivating region (CR3, Figure 1A, Construct 3) were also generated for HPV16 E7 (amino acids 38-98) and HPV1a E7 (amino acids 39-93). All HPV16 E7 proteins, HPV1a E7(1-93) and HPV1a E7 (16-93) constructs express soluble proteins at 37°C and are purified to homogeneity through a combination of anion exchange (Q-sepharose), separation based on hydrophobicity (Phenyl sepharose or ammonium sulfate precipitation) and gel filtration (Superdex-200). In contrast, HPV1a E7 (39-93) protein is insoluble when expressed at 37^{oC} but is refolded and purified to homogeneity with gel filtration (s200). All E7 constructs elute from the gel filtration columns in a single peak in the form of a multimer. Analytical ultracentrifugation sedimentation equilibrium experiments of E7 constructs indicate that this protein exists primarily as a dimer (Table 1).

In addition to the E7 constructs, comparable Adenovirus 5 E1A constructs containing CR1-CR3 (amino acids 36-189), CR2-CR3 (amino acids 114-189) and CR3 (135-189) were generated from Adenovirus 5 E1A cDNA that was obtained from Dr. Ricciardi (Figure 1A, Constructs 4-6). These constructs are expressed in bacteria at 37°C using the same system as described for E7. E1A(36-189) is purified with a combination of anion exchange (Q sepharose), dye affinity chromatography (Reactive Red and Blue sepharose) and gel filtration on a Superdex-200 gel filtration column. However, gel filtration indicates that the E1A(36-189) protein exists as several differently sized multimers that appear to be susceptible to degradation. E1A(114-189) and E1A (135-189) are purified to homogeneity by anion exchange (Q sepharose), ammonium sulfate precipitation and gel filtration.

The retinoblastoma tumor suppressor protein (pRB) contains two domains that are required for minimal viral oncoprotein interaction (domain A and domain B, Figure 1B). These two domains are referred to as the small pocket of pRB. Several pRB constructs containing these domains were subcloned into pRSET A for bacterial expression (Figure 1B, Constructs 1-5). The cDNA for full length pRB was obtained from Dr Ricciardi. All pRB constructs are bacterially expressed in BL21(DE3) cells. These constructs produce soluble proteins when induced at 15°C overnight. The soluble 6x histidine tagged pRB proteins are purified with a combination of affinity chromatography (Ni-NTA agarose) and gel filtration on a Superdex-200 FPLC column. Each of these proteins elute from the gel filtration column in one peak consistent with the molecular weight of monomeric protein.

A bacterial coexpression system has been developed to conveniently prepare suitable pRB/viral oncoprotein complexes (Johnston, KJ, Clements, A, *et. al., manuscript submitted*). With this system, several pRB constructs have been subcloned into a modified version of the pMR103 expression vector (Figure 1B, Constructs 2-5). The kanamycin resistant pMR103 expression vector

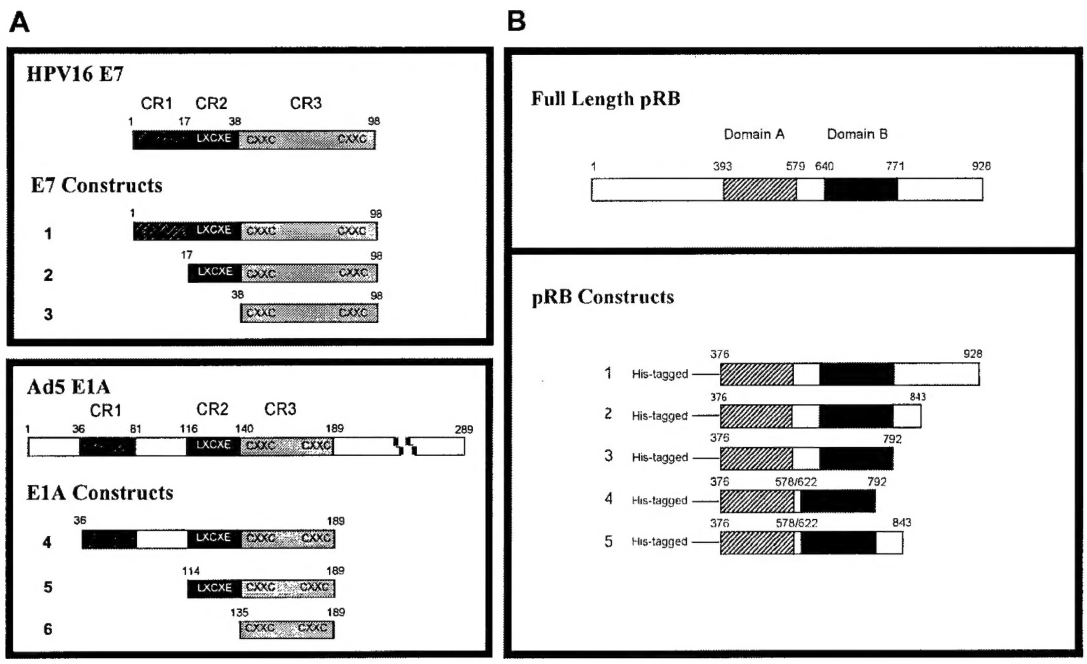


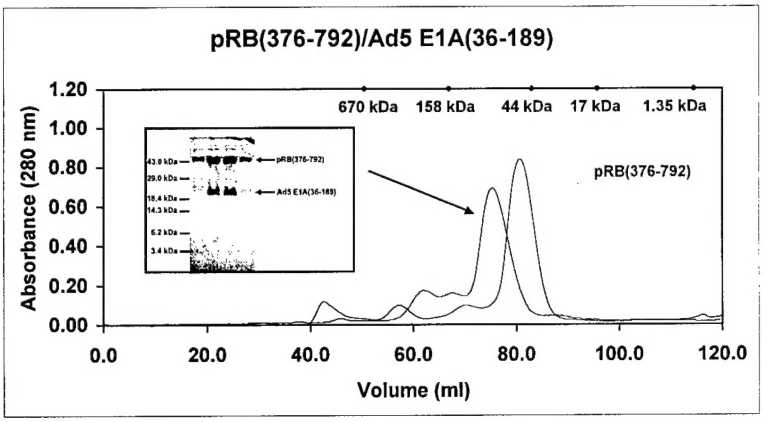
Figure 1. Schematic representation of pRB and viral oncoprotein constructs. (A) The viral oncoproteins contain three conserved regions (CR1,CR2 and CR3). The constructs of HPV16 E7 and Adenovirus 5 E1A are shown schematically. Similar constructs were made for HPV1a E7 as described in the text. CR2 contains the minimal pRB binding region LXCXE (were X represents any amino acid). CR3 contains a Zn^{2+} binding region comprised of two CXXC motifs separated by a linker. The CR3 region of E7 is necessary for pRB inactivation. (B) pRB is a 928 amino acid protein that contains two domains (A and B) which are necessary for viral oncoprotein interaction. Five 6Xhistidine tagged pRB constructs are shown. All constructs contain domain A and domain B. Construct 1 extends to the C-terminus of the protein. Construct 2 and construct 5 extend to the 2nd E7(CR3 region) binding site (amino acids 792-843)[6]. Construct 3 and construct 4 extend to the end of domain B. Construct 4 and construct 5 have deletions from 579 to 621 in the flexible linker region of pRB[4].

is used for pRB coexpression with viral oncoproteins from the ampicillin resistant pRSET vector. pRB and viral oncoproteins are coexpressed in bacteria at 15°C and purified with affinity chromatography (Ni-NTA agarose) and gel filtration (Superdex-200). Several of the following pRB-viral oncoprotein complexes have been purified: (pRB(376-792)/E1A(36-189), pRB(376-792)/E1A(114-189), pRB(376-792)/HPV16 E7(1-98), pRB(376-792)/HPV16 E7(17-98), and pRB(376-843ΔL)/HPV16 E7(17-98)). All of the above purified complexes have been utilized for crystallization trials with several different factorial screens. To date, these complexes have resisted crystallization.

In order to gain insight into pRB function, the oligomerization states of purified pRB, HPV16 E7, Ad5 E1A and pRB/viral oncoprotein complexes were characterized in solution using sedimentation equilibrium experiments with a Beckman XL-I analytical ultracentrifuge. All experiments were performed at 4° C and at multiple speeds and/or concentrations. After sedimentation equilibrium was reached, data plots from several scans were analyzed simultaneously using the program NONLIN. Gel filtration and analytical ultracentrifugation data for the purified pRB(376-792)/Ad5 E1A(36-189) complex and for the purified pRB(376-792)/HPV16 E7(17-98) complex are shown in Figure 2-3. Gel filtration data indicate that the pRB(376-792)/Ad5 E1A(36-189) complex elutes from the column during one peak at a molecular weight that is consistent with a 1:1 stoichiometric complex (Figure 2A). Sedimentation equilibrium experiments indicate that this complex is in a reversibly associating process. The best model for the data indicate that complexes of 1:1 stoichiometry and 2:2

Figure 2

A.



B.

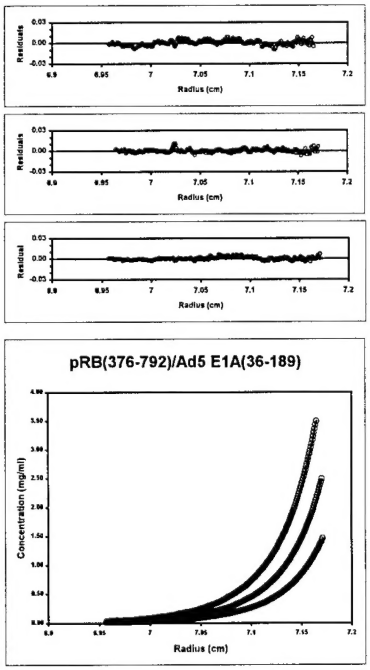
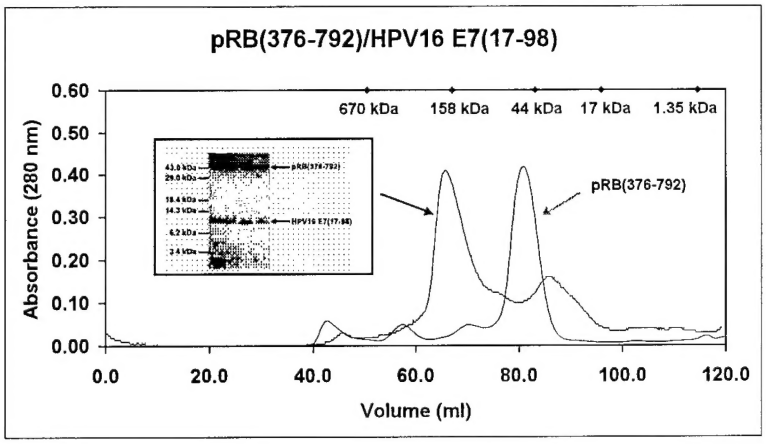


Figure 2. The purified recombinant pRB(376-792)/Ad5 E1A(36-189) complex exists primarily with 1:1 stoichiometry. A. Size exclusion chromatography of the recombinant pRB(376-792)/Ad5 E1A(36-189) complex. An 18% Coomassie-blue stained SDS-PAGE analysis of four peak fractions reveal that coexpressed recombinant pRB(376-792) and Ad5 E1A(36-189) coelute from a Pharmacia superdex 200 preparative size exclusion column between elution points of protein standards with molecular weights of the 44kDa and 158kDa. A 1:1 stoichiometric complex consisting of pRB(376-792) and Ad5 E1A(36-189) has a calculated molecular weight of 67132.2 daltons based on the protein sequences. The grey peak represents the elution point of pRB(376-792) alone (the peak height is scaled down to facilitate peak elution point comparison). The peak elution points of protein standards are represented by the ♦ symbols. B. Sedimentation equilibrium analysis of the pRB(376-792)/Ad5 E1A(36-189) complex at multiple protein concentrations. This single speed sedimentation equilibrium experiment was performed at 18500 rpm and 4°C. Fringe displacement data from three different loading concentrations were analyzed simultaneously with a model describing complexes of 1:1 and 2:2 stoichiometry with individual dissociation constants ranging from 91.4μM to 211.6μM using NONLIN. The bottom panel illustrates the calculated fits as continuous lines. Data points are for all scans are represented as O. The top three panels represent residuals for the calculated fits shown from the highest protein concentration to the lowest protein concentration.

Figure 3

A.



B.

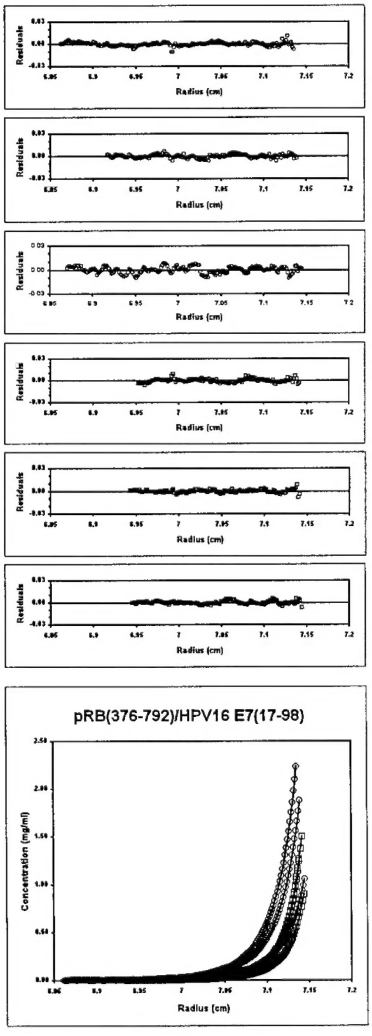


Figure 3. The purified recombinant pRB(376-792)/HPV 16 E7(17-98) complex exists primarily with 2:2 stoichiometry. A. Size exclusion chromatography of the recombinant pRB(376-792)/HPV16 E7(17-98) complex. An 18% Coomassie-blue stained SDS-PAGE analysis of four peak fractions reveal that coexpressed recombinant pRB(376-792) and HPV16 E7(17-98) coelute from a Pharmacia superdex 200 preparative size exclusion column at approximately the elution point of the 158kDa protein standard. A 1:1 stoichiometric complex consisting of pRB(376-792) and HPV16 E7(17-98) has a calculated molecular weight of 59297.6 daltons based on the protein sequences. The grey peak represents the elution point of pRB(376-792) alone (the peak height is scaled down to facilitate peak elution point comparison). The peak elution points of protein standards are represented by the ♦ symbols. B. Sedimentation equilibrium analysis of the purified pRB(376-792)/HPV16 E7(17-98) complex. Two separate 4°C single speed sedimentation equilibrium experiments were performed at 18000rpm and at 20000rpm. Fringe displacement data from three different concentrations at both speeds were analyzed simultaneously with a model describing complexes of 1:1, 2:2 and 4:4 stoichiometry using NONLIN. This fit was significantly better than a model describing complexes of 1:1, 2:2 and 3:3 stoichiometry. The individual dissociation constants for 1:1 - 2:2 stoichiometry equilibrium ranged from 0.7μM to 1.4μM. The individual dissociation constants for 2:2 - 4:4 stoichiometry equilibrium ranged from 11.9μM to 78.4μM. The bottom panel illustrates the calculated fits as continuous lines. Data points are represented as ○ for the experiment at 18000rpm and □for the experiment at 20000rpm. The top three panels represent the residuals from the fits at 18000rpm and are shown from the highest protein concentration to the lowest protein concentration. The next three panels represent the residuals from the fits at 20000rpm and are shown from the highest protein concentration to the lowest protein concentration.

stoichiometry associate reversibly in solution with an apparent dissociation constant (K_d) in the range of 91.4 μM to 211.6 μM (Figure 2B and Table 1). This fit is significantly better than models estimating 2:1 or 1:2 Ad5 E1A/pRB stoichiometry. In contrast, gel filtration data of the pRB(376-792)/HPV16 E7(17-98) complex indicate that the complex elutes from the column during one peak at a molecular weight that is of greater stoichiometry than 1:1 (Figure 3A). Several models could be used to describe the sedimentation equilibrium results of this pRB(376-792)/HPV16 E7(17-98) complex : a model fitting for 1:1-2:2-4:4 molar stoichiometry (Figure 3B and Table 1), a model fitting for 2:1-4:2-8:4 molar stoichiometry and a model fitting for 2:1-4:2-6:3 stoichiometry of HPV16 E7/pRB. Although there was not a statistical difference in the quality of fits for these three models, the first fit had the most randomly distributed residuals of all three models. In all cases, the majority of the complex in solution had greater than a 1:1 stoichiometry of pRB/HPV16 E7. The apparent dissociation constants of all proteins tested in solution are summarized in Table 1.

Table 1. Apparent Dissociation Constants (K_d) of pRB(376-792), HPV16 E7(1-98), Ad5 E1A(114-189), pRB(376-792)/HPV16 E7(17-98) and pRB(376-792)/Ad5 E1A(36-189)

Protein	$K_{d(monomer-dimer)}$	$K_{d(dimer-tetramer)}$
pRB (376-792)	0.60-2.06 mM	N/A
HPV16 E7 (1-98)	0.73-6.71 μM	251-909 μM
Ad5 E1A (114-189)	146 –238 μM	N/A

pRB/ Viral Oncoprotein Complex	$K_{d(1:1-2:2)}$	$K_{d(2:2-4:4)}$
pRB (376-792)/ HPV16 E7(17-98)	0.24-1.38 μM	11.9-78.4 μM
pRB (376-792)/ Ad5 E1A (36-189)	91.4-212 μM	N/A

From these sedimentation equilibrium experiments, it is concluded that pRB(376-792) exists primarily as a monomer. Ad5 E1A (114-189) and HPV16 E7 (1-98) participate in reversibly associating processes with significantly different dissociation constants. HPV16 E7 (1-98) exists in monomer-dimer-tetramer equilibrium at the concentrations tested. Monomer-dimer equilibrium was also detected with Ad5 E1A (114-189). However, the apparent $K_{d(monomer-dimer)}$ of Ad5 E1A is approximately 100-fold lower than the apparent $K_{d(monomer-dimer)}$ of HPV16 E7, demonstrating that Ad5 E1A (114-189) exists primarily as a monomer and HPV16 E7 exists primarily as a dimer in solution. The pRB/viral oncoprotein complexes also participate in reversibly associating processes. The apparent $K_{d(1:1-2:2)}$ of pRB(376-792)/Ad5 E1A(36-189) is comparable to the $K_{d(monomer-dimer)}$ of Ad5 E1A (114-189), suggesting that pRB/Ad5 E1A oligomerization is mediated through Ad5 E1A. Several models for pRB(376-792)/HPV16 E7(17-98) suggest that the stoichiometry of the complex is greater than 1:1. Thus, it appears that viral oncoprotein oligomerization is not inhibited by pRB-binding.

Although pRB/viral oncoprotein complexes have resisted crystallization, crystals of the HPV1a E7(39-93) protein have been obtained and tested for diffraction (Figure 4). These E7 crystals diffracted to approximately 2.8Å at the Brookhaven synchrotron beamline X4A. However, the diffraction spots from the crystal could not be processed and indicate that the crystals are highly mosaic. These HPV1a E7 crystals are currently being refined in order to obtain higher quality diffracting crystals. In addition to E7 crystallization, an excellent NMR spectrum of Adenovirus 5 E1A (135-198) CR3 region has been obtained recently. The NOE and chemical shift data in the

Figure 4. Crystals of HPV1a E7(39-93) protein. Purified HPV1a E7(39-93) protein was crystallized by a hanging drop vapor diffusion method. The reservoir crystallization condition contains 0.5-1.5 M NaCl, 20-30 % Ethanol and 0.1M Hepes, pH 7.5. Crystals tend to grow as plates with the average dimensions of 200μm X 200μm X 20μm.

2D-NOESY spectrum of E1A are consistent with the presence of a mixed β-strand, α-helix structure (Figure 5). Therefore, the NMR structure of the E1A CR3 region is being pursued as well.

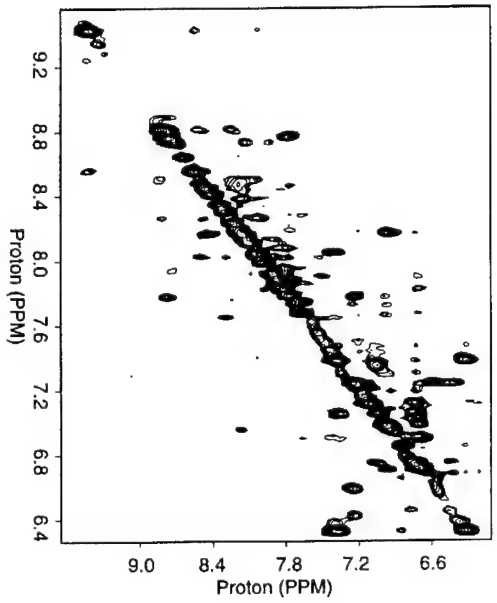


Figure 5. Preliminary 2D-NOESY spectrum for the Adenovirus 5 E1A CR3 region. Adenovirus 5 E1A(135-189) was used in this study. The NOE and chemical shift data for this domain are consistent with the presence of a mixed β-sheet, α-helix structure. The line widths and relaxation times of protons are consistent with a monomeric Adenovirus 5 E1A species.

PCAF transcriptional coactivator studies

A DNA construct that encoded for amino acids 493 to 658 of PCAF (plus N-terminal met-lys) was subcloned into the pRSET-A vector for bacterial expression. After overexpression of PCAF at 15°c for 12 hours, the majority of the recombinant protein was found in the soluble cell extract. Soluble p/CAF protein was purified by a combination of cation exchange chromatography (SP-sepharose), affinity chromatography (Coenzyme A-agarose) and size exclusion chromatography (Superdex 200). The monomeric protein was then concentrated to approximately 20-40 mg/ml, flash frozen, and stored at -70°C.

For protein crystallization, 10mg/ml of the PCAF protein was mixed with 2-fold molar excess of Na-acetyl coenzyme A. Crystals were grown by the hanging drop vapor diffusion method at room temperature. The crystallization mixture contained 1.3-1.6M Li_2SO_4 and 0.1M Tris-Cl (pH 8.5). Rod-shaped crystals generally appeared after 2 to 3 weeks with average cell dimensions of 0.2mm X 0.08mm X 0.08mm. Crystals were then slowly and sequentially transferred into a cryoprotectant solution

containing 1.5M Li_2SO_4, 0.1M Tris-Cl (pH 8.5), and 15% Ethanol. The crystals were then flash frozen in 15% ethanol with liquid propane for data collection. A native data set was collected on Beamline X4-A (λ=1.0009Å) at the National Synchrotron Light Source at Brookhaven National Laboratory. The data was collected in 1° oscillations using a Raxis-IV area detector. The programs Denzo and Scalepack[13] were used to process and scale the data.

Two solutions for the PCAF-coenzyme A complex were obtained at 10.0 to 4.0Å by molecular replacement using the coordinates of a partially refined model (R=30.1%, R_{free}=34.1%) of *apo Tetrahymena* GCN5 with program AMORE[14]. Prior to initial rigid body refinement, a randomly chosen 10% of the total number of reflections was designated as a test data set and all residues in the model that were not identical to PCAF residues were alanized. The initial electron density maps generated with Fourier coefficients $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ showed clear side chain density for most of the PCAF specific residues. The alanized residues with p/CAF-specific side chain $|F_o|-|F_c|$ electron density were replaced by p/CAF-specific residues using the program O. After one round of simulated annealing from 8.0 to 3.0A, $|F_o|-|F_c|$ electron density maps showed strong peaks for the pantothenic acid and the pyrophosphates of the 3'phosphate ADP moiety in Coenzyme A. Refinement proceeded by multiple rounds of positional refinement, simulated annealing, and torsion-angle dynamics with periodic model building in O. Refinement was carried out in resolution steps of 3.0, 2.7, 2.5, and 2.3A using the program X-PLOR 3.8[15] and CNS-SOLVE[16]. At the final stages of refinement, a bulk solvent correction was applied using data from 20.0-2.3A and tightly constrained B-factor refinement was performed using the program CNS-SOLVE. Ordered water molecules were built into strong $|F_o|-|F_c|$ peaks and only retained if possible H-bond partners could be located and if they refined to a reasonable B factor. Protein A in the asymmetric unit resulted in a model containing amino acids 493-653 plus an N-terminal lysine. Protein B in the asymmetric unit resulted in a model containing 493-652 plus the N-terminal lysine. For protein B only, density was not observed for solvent exposed residue side chains 503, 505, 625, 626, 627, 631, 636 and were therefore modeled as alanines. The final model of each protein-CoA complex has good geometry (Table 2) with none of the non-glycine residues lying in disallowed regions of the Ramachandran plot.

The 2.3Å crystal structure of the PCAF protein acetyltransferase domain reveals an α/β globular fold that contains a central protein core which sits at the base of a pronounced cleft that is formed by the N- and C-terminal protein segments (Figure 5). The protein core at the base of this cleft makes extensive contacts with the pantetheine arm of coenzyme A, marking the active site of the enzyme. of this structure with extensive mutagenesis data for PCAF and for the homologous yeast GCN5 protein implicates this cleft and the N- and C-terminal segments to play an important role in histone or p53 substrate binding. Inspection of this mutationally sensitive region suggests that a glutamate residue within the protein core plays a catalytic role for protein acetylation. From this crystallographic study, a catalytic mechanism for the

Table 2. Crystallographic Data and Refinement Statistics

Crystal Parameters		**Data Collection Statistics**	
Unit Cell Dimensions		**Resolution Range**	**20.0-2.3 Å**
a=97.00 Å , b=97.00 Å, c=77.85 Å		**Total Reflections**	**94,731**
α=90.00°, β=90.00°, γ=120.00°		**Unique Reflections**	**17,943**
		R_{sym} (%)	**4.0 (15.5)**
Space Group	**$P6_4$**	***I*/sigma (*I*)**	**18.0 (4.8)**
Asymmetric Unit	**2 molecules**	**Completeness (%)**	**96.5 (99.8)**
Refinement Statistics			
Resolution Range	**20.0-2.3 Å**	**R.m.s. Values**	
***I*/σ cutoff**	**0.0**	**Bond length (Å)**	**0.007**
Final Model		**Bond angles (°)**	**1.89**
Protein atoms	**2606**	**NCS molecules (Å)**	**1.38**
Water atoms	**109**	**B-factors (Å²)**	**1.64**
CoA atoms	**96**	**Average B-factors (Å²)**	
R-factors		**Protein (A/B)[a]**	**31.5/40.7**
$R_{working}$	**22.3%**	**Water**	**39.0**
R_{free}	**26.8%**	**CoA (A/B)[a]**	**39.4/52.2**

R-factor: $R_{working} = \Sigma \| |F_o| - |F_c| \| / \Sigma |F_o|$; $R_{free} = \Sigma_T \| |F_o| - |F_c| \| / \Sigma_T |F_o|$, where T is a test data set of 10% of the total reflections randomly chosen and set aside before refinement.

[a]A and B refer to complexes A and B in the asymmetric unit cell. The numbers in parentheses are for the highest resolution bins.

acetylation of histones and of p53 is proposed. In order to gain insight into substrate specificity, x-ray crystallographic studies are being performed on PCAF and on the homologous *Tetrahymena* GCN5 protein with substrate peptides derived from histones and from p53.

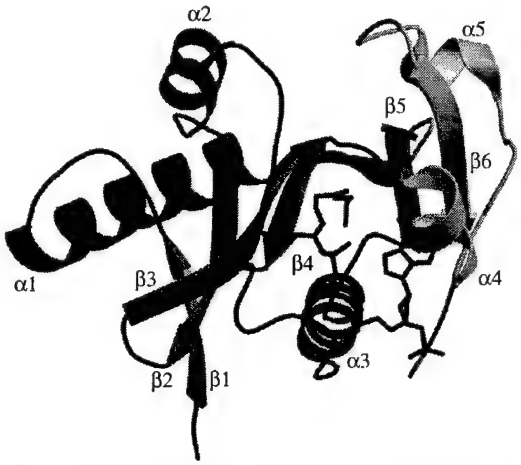


Figure 5. Structure of the PCAF-coenzyme A complex.

In summary, this study has demonstrated that HPV E7 exists predominantly as a dimer, while comparable constructs of Adenovirus 5 E1A are primarily monomeric. Apparent dissociation constants were determined for these proteins using sedimentation equilibrium experiments. Bacterial coexpression can be utilized to form stable pRB/viral oncoprotein complexes. Ad5 E1A/ pRB and HPV16 E7/pRB complexes reversibly self-associate and dissociation constants were determined for these complexes using sedimentation equilibrium experiments. Ad5 E1A/pRB oligomerization has a comparable dissociation constant to E1A monomer-dimer equilibrium, suggesting that Ad5 E1A/pRB oligomerization is mediated through E1A. Unlike pRB/E1A, pRB/E7 exists primarily with a stoichiometry greater than 1:1. pRB/viral oncoprotein complexes have resisted crystallization to date. The HPV1a E7(39-93) crystals are being refined to obtain high quality diffracting crystals. Structural studies using NMR are currently being performed with the Adenovirus 5 E1A(135-189) CR3 region. Additionally, the crystal structure of the human PCAF acetyltransferase domain has been solved to 2.3Å and provides tremendous insight into the mechanism of histone acetylation and p53 activation. In order to gain insight into PCAF substrate specificity, current structural studies are being performed with PCAF and the homologous *Tetrahymena* GCN5 protein with substrate histone-derived and p53-derived peptides.

KEY RESEARCH ACCOMPLISHMENTS

- Several constructs of recombinant DNA viral oncoproteins, HPV16 E7 and Ad5 E1A, were purified to homogeneity.

- Despite the considerable sequence and functional homology of HPV16 E7 and Ad5 E1A , sedimentation equilibrium experiments reveal that dissociation constants of these proteins differ significantly.

- Several constructs of recombinant pRB were purified to homogeneity.

- Sedimentation equilibrium experiments reveal that pRB(376-792) is monomeric in solution.

- A dual vector bacterial coexpression system was developed to make significant quantities of purified pRB/viral oncoprotein complexes.

- Sedimentation equilibrium experiments revealed that HPV16 E7 and Ad5 E1A oligomerization is not inhibited by pRB-binding.

- The crystal structure of the PCAF transcriptional coactivator bound to coenzyme A reveals the molecular mechanism of histone acetylation.

REPORTABLE OUTCOMES

Clements A, Rojas JR, Trievel RC, Wang L, Berger SL, Marmorstein R. Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. The EMBO J1999;18(13):3521-3532.

Johnston K*, Clements A*, Venkataramani RN, Trievel RC, Marmorstein R. Coexpression of proteins in bacteria using T7-based expression plasmids: expression of heteromeric cell-cycle and transcriptional regulatory complexes. Manuscript submitted for publication.

*The first two authors contributed equally to this work.

CONCLUSIONS

Despite considerable sequence and functional homology of DNA viral oncoproteins, Ad5 E1A and HPV16 E7, these proteins have different oligomerization properties. HPV16 E7 exists primarily as a dimer and Ad5 E1A is primarily monomeric. Apparent dissociation constants were determined for these proteins utilizing analytical ultracentrifugation experiments. A bacterial coexpression system was utilized to coexpress and copurify pRB/viral oncoprotein complexes. This provided the means necessary to make significant quantities of highly pure complexes for biophysical characterization and for crystallization trials. Although crystallization of the pRB/viral oncoproteins was problematic, these complexes were well characterized biophysically in this study. Specifically, it was determined that pRB-binding does not inhibit HPV16 E7 or Ad5 E1A oligomerization. pRB/viral oncoprotein stoichiometry was characterized as well. This research has provided a model to develop large quantities of purified recombinant cell cycle regulatory complexes for biophysical, biochemical and structural studies. In addition, these sedimentation equilibrium experiments are the first example of pRB/viral oncoprotein oligomerization studies in solution. Additionally, insight into the mechanism of histone acetylation and of p53 acetylation was achieved by X-ray crystallographic studies of the PCAF transcriptional coactivator bound to coenzyme A.

REFERENCES

1. Porter-Jordan K., Lippman M.E. (1994) Overview of the biologic markers of breast cancer. *Hematol Oncol Clin North Am* **8**(1),73-100.
2. Hunter T., Pines, J. (1994) Cyclins and cancer II: cyclin D and CDK inhibiyors come of age. *Cell.* **79**, 573-582.
3. Bartkova J., *et al.* (1994) Cyclin D1 protein expression and function in human brest cancer. *Int. J. Cancer.* **57**, 353-361.
4. Lee, J.-O., *et al.* (1998) Structure of the retinoblastoma tumour –suppressor pocket domain bound to a peptide from HPV E7. *Nature*. **391**, 859-865.
5. Jones, R.E., *et. al* . (1990) Identification of the HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. *J. Biol. Chem.* **265**, 12782-12785.
6. Patrucjm D.R., *et.al.* (1994) Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein. *J. Biol.Chem.* **269**, 6842-6850.
7. Ikeda, M.-A., *et al.* (1993) Identification of distinct roles for separate E1A domains in the disruption of E2F complexes. *Mol. Cell. Biol.* **13**(11), 7029-7035.
8. Amundson S.A., *et al.* (1998) Roles for p53 in growth arrest and apoptosis: putting on the brakes after genotoxic stress. *Oncogene* **17**(25),3287-3299.
9. Elledge R.M., Allred D.C. (1998) Prognostic and predictive value of p53 and p21 in breast cancer. *Breast Cancer Res. Treat.* **52**(1-3),79-98
10. Scolnick, D., *et al.* (1997) CREB-binding protein and p300/CBP-associated factor are transcriptional coactivators of the p53 tumor suppreesor protein. *Canc. Res.* **57**, 3693-3696.
11. Liu, L., *et al.* (1999) p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. *Mol. Cell. Biol.* **19**, 1202-1209.
12. Charkravarti, D. *et al.* (1999) A viral mechanism for inhibition of p300 and PCAF acetyltransferase activity. *Cell.* **96**(3), 393-403.
13. Otwinowski, Z. (1993) Oscillation data reduction program. In Sawyer, L., et. al. (eds), Proceedingsl of the CCP4 study weekend: data collection andprocessing. SERC Daresbury Laboratory, Warrington, UK, 56-62.
14. Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157-163.
15. Brunger, A.T., (1992) X-PLOR 3.1 : A system for X-ray crystallography and NMR. Yale University Press, New Haven, CT.
16. Brunger, A.T., *et al.* (1998) Crystallography and NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905-921.

The EMBO Journal Vol.18 No.13 pp.3521–3532, 1999

Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A

Adrienne Clements[1,2], Jeannie R.Rojas[1,3], Raymond C.Trievel[1,2], Lian Wang[1,2], Shelley L.Berger[1] and Ronen Marmorstein[1,2,3,4]

[1]The Wistar Institute, [3]Department of Chemistry and the [2]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA

[4]Corresponding author
e-mail: marmor@wistar.upenn.edu

The human p300/CBP-associating factor, PCAF, mediates transcriptional activation through its ability to acetylate nucleosomal histone substrates as well as transcriptional activators such as p53. We have determined the 2.3 Å crystal structure of the histone acetyltransferase (HAT) domain of PCAF bound to coenzyme A. The structure reveals a central protein core associated with coenzyme A binding and a pronounced cleft that sits over the protein core and is flanked on opposite sides by the N- and C-terminal protein segments. A correlation of the structure with the extensive mutagenesis data for PCAF and the homologous yeast GCN5 protein implicates the cleft and the N- and C-terminal protein segments as playing an important role in histone substrate binding, and a glutamate residue in the protein core as playing an essential catalytic role. A structural comparison with the coenzyme-bound forms of the related *N*-acetyltransferases, HAT1 (yeast histone acetyltransferase 1) and SmAAT (*Serratia marcescens* aminoglycoside 3-*N*-acetyltransferase), suggests the mode of substrate binding and catalysis by these enzymes and establishes a paradigm for understanding the structure–function relationships of other enzymes that acetylate histones and transcriptional regulators to promote activated transcription.

Keywords: acetyltransferase/coactivator HAT/p300/CBP-associating factor

Introduction

The PCAF (p300/CBP-associating factor) transcriptional coactivator was identified initially through its ability to interact with p300/CBP for the transcriptional activation of many genes, and to counteract the ability of the adenoviral E1A oncoprotein to inhibit p300/CBP-mediated transcriptional activation (Yang *et al.*, 1996). The same study showed that PCAF contains intrinsic histone acetyltransferase activity, a property previously demonstrated for the GCN5 transcriptional coactivator (Marcus *et al.*, 1994; Brownell *et al.*, 1996), and is correlated with transcriptional activation (Brownell and Allis, 1996; Wolffe and Pruss, 1996; Grunstein, 1997). More recently, PCAF has also been shown to interact with the DNA-binding domain of nuclear receptors such as RXR/RAR, independent of p300/CBP binding, to promote retinoid-responsive transcriptional activation (Blanco *et al.*, 1998), and has been shown to interact directly with E1A resulting in an inhibition of its intrinsic histone acetyltransferase activity and its ability to mediate transcriptional activation (Reid *et al.*, 1998; Chakravarti *et al.*, 1999).

Analysis of the primary sequence of the 832 residue PCAF protein reveals that it contains a C-terminal bromodomain (within residues 725–819), a central histone acetyltransferase (HAT) domain (within residues 493–653) highly homologous to the GCN5 transciptional coactivator [from *Tetrahymena* (Brownell *et al.*, 1996) and from yeast (Marcus *et al.*, 1994)] and a structurally divergent N-terminal region (Yang *et al.*, 1996). More recently, Roth and colleagues have shown that the N-terminal region of PCAF shares homology with the predominant form of mammalian GCN5 (Xu *et al.*, 1998). Functional characterization of the N-terminal segment of PCAF shows that it contains an interaction surface for p300/CBP (Yang *et al.*, 1996; Xu *et al.*, 1998), other transcriptional activators (Currie, 1998; Krumm *et al.*, 1998) and E1A (Chakravarti *et al.*, 1999), and is required for nucleosomal acetylation mediated by the PCAF HAT domain (Yang *et al.*, 1996).

The HAT domain of PCAF has been analyzed extensively at the amino acid and functional levels. The HAT domain of PCAF shares a high degree of sequence homology with GCN5 from various species (GCN5/PCAF subfamily of histones acetyltransferases) (Marcus *et al.*, 1994; Brownell *et al.*, 1996; Candau *et al.*, 1996; Smith *et al.*, 1998a) and has functional homology with other transcriptional coactivators that harbor HAT activity including yeast ESA1 (Smith *et al.*, 1998b), and human CBP/300 (Ogryzko *et al.*, 1996), $TAF_{II}250$ (Mizzen *et al.*, 1996), Tip60 (Yamamoto and Horikoshi, 1997), ACTR (Chen *et al.*, 1997) and SRC-1 (Spencer *et al.*, 1997). More recently, detailed sequence analysis has revealed that the HAT domain of PCAF shares limited sequence homology with a biologically diverse family of GCN5-related *N*-acetyltransferases (GNATs) within three relatively small motifs (15–33 residues) called A, D and B (Neuwald and Landsman, 1997).

In vivo, PCAF has been shown to function in the context of a large multisubunit protein complex with >20 distinct polypeptides including several of the TATA-binding protein (TBP)-associated factors (TAFs) and human counterparts to the yeast ADA2, ADA3 and SPT3 proteins (Ogryzko *et al.*, 1998). The histone substrate specificity of PCAF has been characterized, showing a strong preference for histone H3 and to a lesser extent histone H4 (Yang *et al.*, 1996; Xu *et al.*, 1998). Interestingly, unlike yeast GCN5 (Kuo *et al.*, 1996, 1998; Wang *et al.*, 1998),

the histone preference of PCAF appears to be similar for both free and nucleosomal histones. Surprisingly, PCAF has also been reported to acetylate non-histone substrates including the basal transcription factors TFIIF and the β-subunit of TFIIE (Imhof *et al.*, 1997). Recently, we have reported that PCAF specifically acetylates Lys320 of the p53 transcriptional activator *in vitro*, resulting in an increased affinity of p53 for DNA (Liu *et al.*, 1999). Correlatively, we find that these same sites are acetylated *in vivo* in response to DNA damage. Most recently, the histone acetyltransferase activity of PCAF towards both nucleosomal histones and p53 has been shown to be inhibited by the direct binding of E1A to its HAT domain (Chakravarti *et al.*, 1999). In order to obtain a detailed view of the mechanism of protein acetylation by PCAF, we have determined the crystal structure of its HAT domain in complex with coenzyme A to a resolution of 2.3 Å.

Results and discussion

Overall structure of the PCAF–coenzyme A complex

The HAT domain of human PCAF (residues 493–658) was overexpressed in *Escherichia coli* and purified to homogeneity using a combination of cation exchange, coenzyme A affinity and gel filtration chromatography. Crystals were obtained containing two protomers per asymmetric unit and the structure was determined by molecular replacement using the unrefined structure of the nascent HAT domain of *Tetrahymena* GCN5 as a search model (J.R.Rojas, R.C.Trievel, Y.Mo, X.Li, J.Zhou, S.L.Berger, C.D.Allis and R.Marmorstein, submitted) (Table I). The two PCAF protomers in the asymmetric unit make modest interprotein interactions and have nearly identical structure, with an r.m.s. deviation between all atoms of 1.38 Å.

PCAF has a βαααββββαβααβ topology and contains a globular fold except for a pronounced cleft along one side of the protein (Figure 1A). It is convenient to think of the core as being formed by two tertiary structural elements near the center of the protein. The first element contains β-strands 2, 3 and 4 aligned in an antiparallel orientation on top of helix α3, while the second element is formed by an adjacent β5-strand–loop–α4-helix. The coenzyme A cofactor is bound between the two elements of the core along one edge of the protein with its labile sulfhydryl pointing into the protein cleft which is flanked on opposite sides by the N- and C-terminal domains of the protein. Within the N-terminal domain, a β-strand forms sheet interactions with the β2-strand of the core, and a helix–turn–helix (α1–turn–α2) sits on one side of the protein above the core. The C-terminal domain contains a helix–loop–strand (α5–loop–β6) which lies opposite the N-terminal domain above the protein core and interacts with the core domain through parallel sheet interactions between β5 and β6.

Mode of coenzyme A binding by PCAF

The coenzyme A cofactor is bound in a cavity formed on the surface of the core region of PCAF and buries over one-half of the coenzyme A accessible surface area and ~520 Å^2 of protein surface area (Figures 1A and 2). It is flanked by the β4–loop–α3 segment that corresponds to motif A of the GNAT proteins on one side and the β5–loop–α4 segment corresponding to motif B of the GNAT proteins on the other side (Figures 1B and 2). Coenzyme A is bound in a bent conformation (Figure 2C) which helps facilitate an extensive set of protein interactions that are mediated predominantly by the pantetheine arm and the pyrophosphate group of coenzyme A (Figure 2A). Strikingly, all but two groups of the 16 member pantheteine arm–pyrophosphate chain are contacted by the protein. All but one of these contacts are mediated through either protein backbone hydrogen bonds or protein side chain van der Waals contacts.

PCAF residues in the GNAT conserved motifs A and B interact extensively with coenzyme A. Specifically, residues 580 and 582–587 in the β4–loop–α3 region of motif A make an extensive set of both direct and water-mediated hydrogen bonds with the pyrophosphate group (Figure 2). Thr587 also makes the only side chain hydrogen bond to the coenzyme, through a pyrophosphate oxygen. The aliphatic side chain of Gln581 and a Cys–Ala–Val sequence (residues 574–576) at the tip of the β4-strand makes an extensive set of van der Waals contacts throughout most of the length of the pantetheine arm. In addition, the backbone residues of Cys574 and Val576 form hydrogen bonds with the pantetheine arm. Residues in the β5–loop–α4 region of GNAT motif B make predominantly van der

Table I. Data and refinement statistics

Crystal parameters	
Unit cell dimensions	a = 97.00 Å, b = 97.00 Å, c = 77.85 Å α = 90.00°, β = 90.00°, γ = 120.00°
Space group	$P6_4$
Asymmetric unit	2 molecules
Data collection statistics	
Resolution range	20.0–2.3 Å
Total reflections	94 731
Unique reflections	17 943
R_{sym}	4.0% (15.5%)
$I/\sigma(I)$	18.0 (4.8)
Completeness	96.5% (99.8%)
Refinement statistics	
Resolution range	20.0–2.3 Å
I/σ cutoff	0.0
Final model	
Protein atoms	2606
Water atoms	109
CoA atoms	96
$R_{working}$	22.3%
R_{free}	26.8%
R.m.s. values	
Bond length (Å)	0.007
Bond angles (°)	1.89
NCS molecules (Å)	1.38
B-factors (Å^2)	1.64
Average B-factors (Å^2)	
Protein (A/B)[a]	31.5/40.7
Coenzyme A (A/B)[a]	39.4/52.2
Water	39.0

$R_{working} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$.
$R_{free} = \Sigma_T||F_o| - |F_c||/\Sigma_T|F_o|$, where T is a test data set of 10% of the total reflections randomly chosen and set aside before refinement.
[a]A and B refer to complexes A and B in the asymmetric unit cell. The numbers in parentheses are for the highest resolution bins.

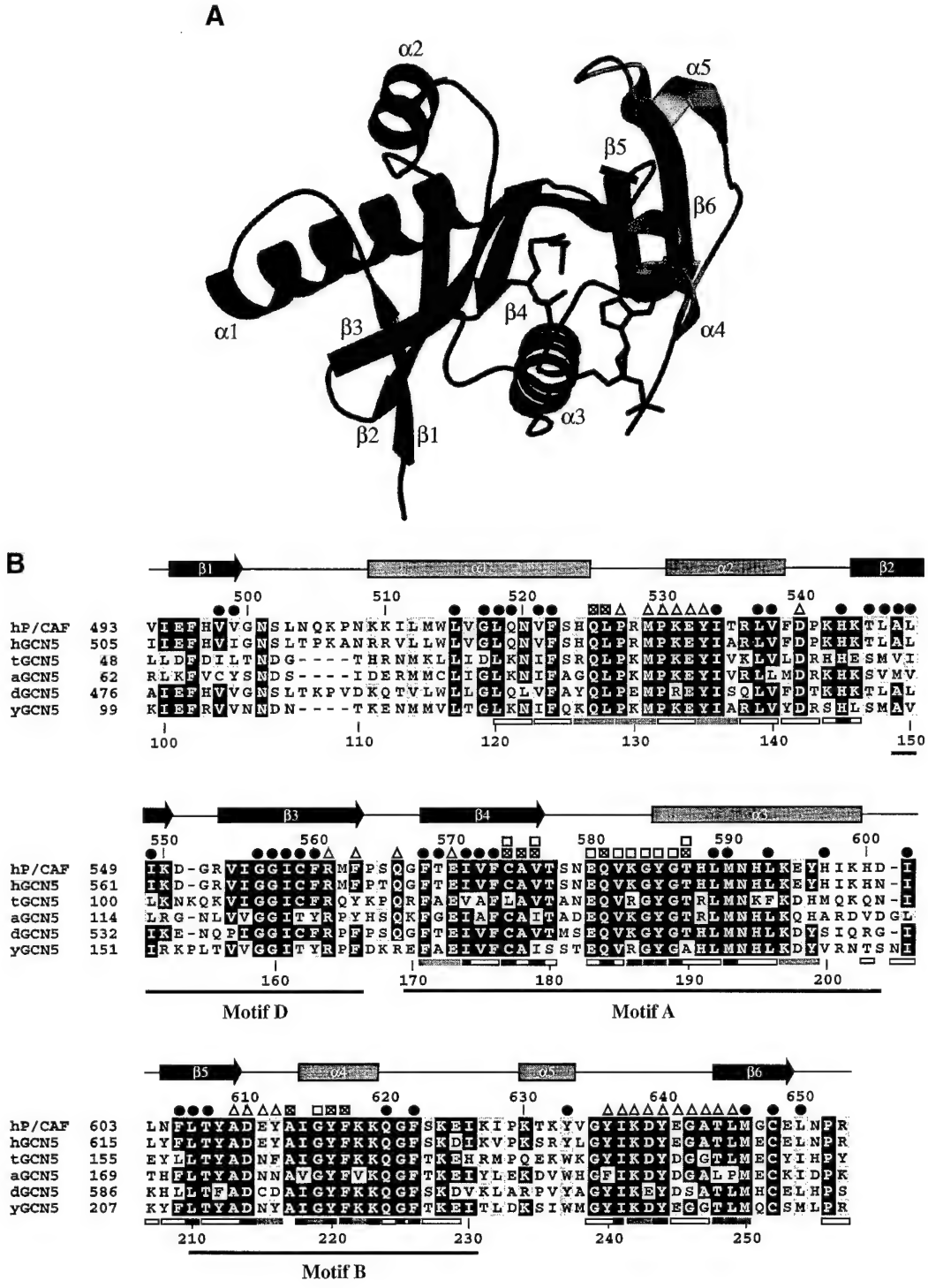


Fig. 1. (**A**) Structure of the PCAF–coenzyme A complex. The four domains of the protein are color-coded, with the two structurally conserved subdomains that make up the core, motifs A–D and motif B′ (based on structural conservation), colored blue and green, respectively. The N- and C-terminal protein segments flanking the core are colored magenta and gold, respectively. Coenzyme A is colored red. (**B**) Sequence alignment of the GCN5/PCAF family of HAT domains. The primary sequence of the HAT domain of human PCAF (hP/CAF) used for the structure determination is shown at the top of the alignment. Sequences from the homologous HAT domains from GCN5 of yeast, *Arabidopsis*, *Drosophila*, human and *Tetrahymena* are aligned (CLUSTAL program) and displayed (BOXSHADE program). Black and gray backgrounds are used to indicate identical and/or conserved residues found in at least 50% of the proteins at a given position, respectively. Secondary structural elements within the HAT domain of PCAF are shown above the sequence alignment. The ● symbol indicates residues that are buried within the core of the protein, the □ symbol indicates residues that contact the coenzyme A cofactor via backbone hydrogen bonds, the ⊠ symbol indicates residues that contact coenzyme A through side chain interactions, and the △ symbol indicates residues that are highly conserved within the GCN5/PCAF family and that are in sufficient proximity to facilitate substrate binding and/or catalysis. Positions of alanine mutations that decrease HAT activity in human PCAF (Martinez-Balbas *et al.*, 1998) and in the homologous yeast GCN5 protein are indicated below the sequence alignment: triple mutations are indicated with gray bars (Wang *et al.*, 1998) and single mutations are indicated with black bars (Kuo *et al.*, 1998). Mutations that have a negligible effect on the HAT activity of yGCN5 are indicated with open rectangles. GNAT motifs D, A and B identified by Neuwald and Landsman (1997) are indicated below the sequence alignment.

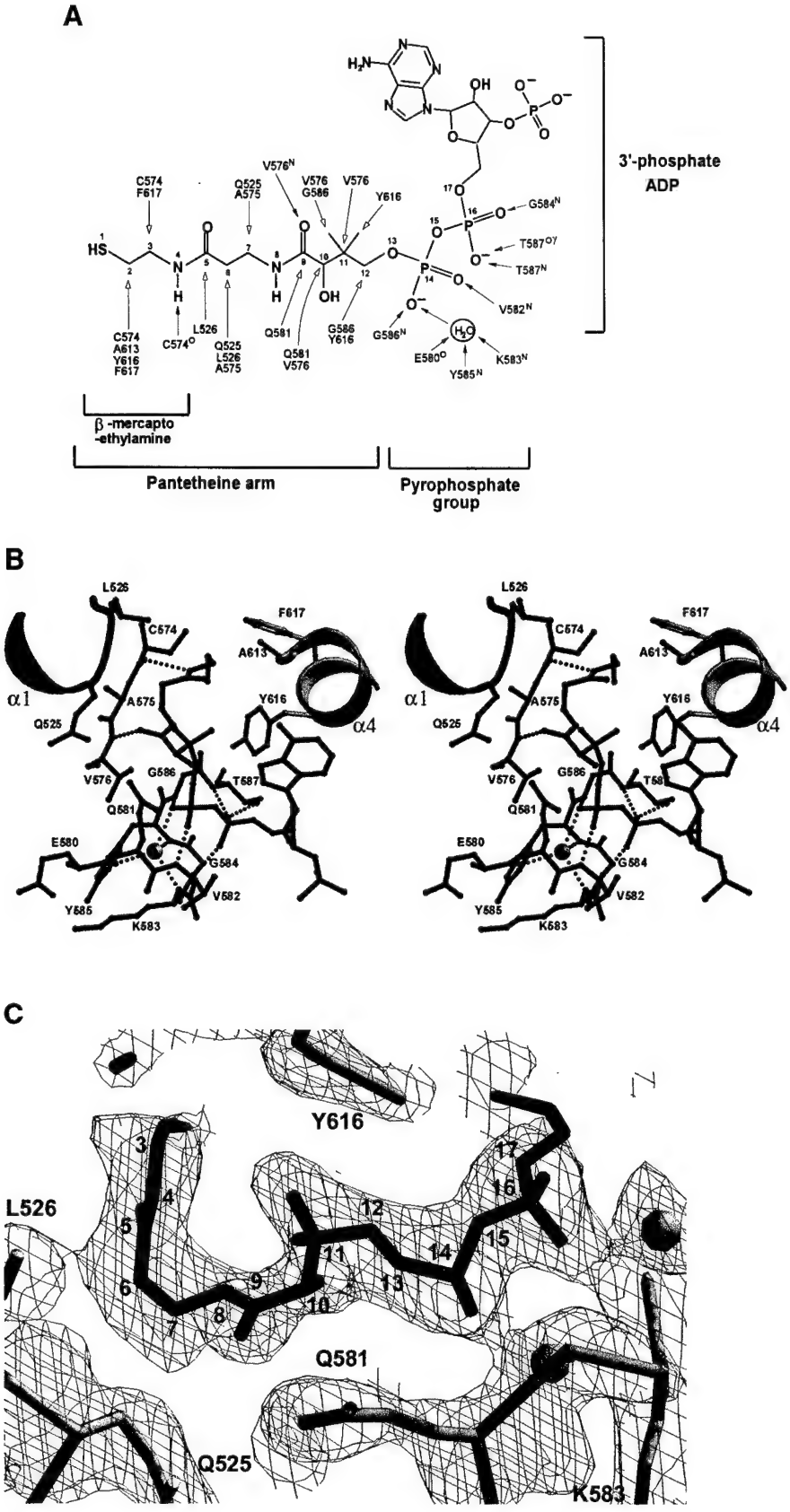


Fig. 2. The coenzyme A-binding site. (**A**) Schematic drawing of PCAF interactions with coenzyme A. Hydrogen bonds are indicated with black arrows, and van der Waals interactions are indicated with white arrows. (**B**) Coenzyme A–protein interactions. Protein residues that make van der Waals contacts and hydrogen bonds (dotted line) are indicated. (**C**) σA-weighted F_o–F_c omit map around the pantetheine arm of the coenzyme A cofactor. The map was generated by omitting residues within a 4.5 Å radius of the cofactor followed by simulated annealing dynamics refinement at a temperature of 1000 K. The map is contoured at 1.5 σ. A portion of coenzyme A is indicated in red and the surrounding protein is indicated in green. The gold spheres indicate ordered water molecules.

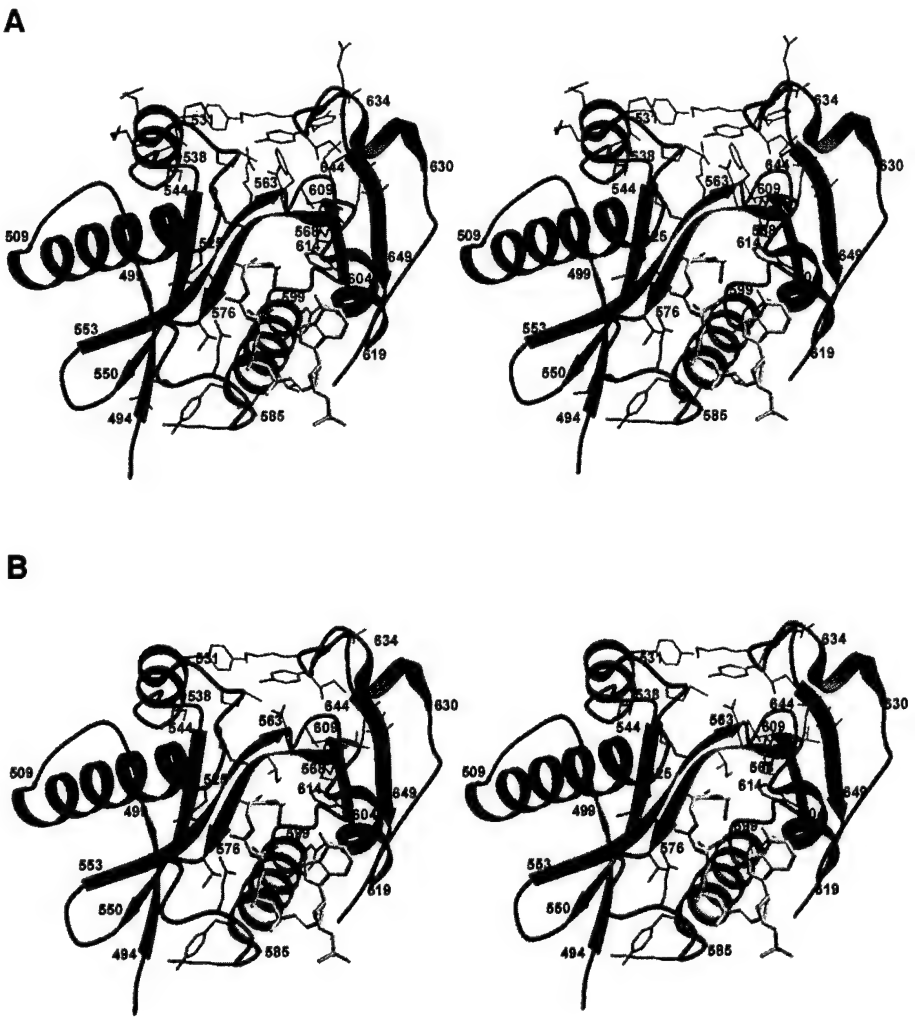


Fig. 3. Functional implications of the PCAF–coenzyme A complex. (**A**) Highly conserved residues within the GCN5/PCAF subfamily of acetyltransferases are mapped onto the PCAF protein. Residues that are associated with coenzyme A interaction are shown in red, residues that are implicated in substrate binding and/or catalysis are shown in green. The remaining strictly conserved residues that are largely buried and presumably important for protein stability are omitted for clarity. Residue numbers at the borders of secondary structural elements are indicated for reference. (**B**) Mutations in human PCAF (Martinez-Balbas *et al.*, 1998) and yeast GCN5 (Kuo *et al.*, 1998; Wang *et al.*, 1998) that decrease HAT activity are mapped onto a schematic representation of the PCAF HAT domain. The color coding is the same as in (A), and residues involved in protein stability are shown in gold.

Waals contacts with the β-mercaptoethylamine segment of the pantetheine arm and thus play a major role in orienting the reactive sulfhydryl atom (atom 1, Figure 2) for acetyl transfer. The protein residues involved are Ala613, Tyr616 and Phe617, while Tyr616 also makes van der Waals contacts with the end of the pantetheine arm near the pyrophosphate group.

Two residues in the non-conserved (within the GNAT family) N-terminal segment of PCAF also interact with coenzyme A. These residues, Gln525 and Leu526, which sit above the core and on one side of the putative substrate-binding cleft, make van der Waals contacts with the pantetheine arm of coenzyme A (Figure 2). The proximity of these residues to the cofactor–substrate junction suggests that they play an important role in substrate-specific binding and/or catalysis. In contrast to the pantetheine arm and pyrophosphate group of coenzyme A, which make extensive protein interactions that are conserved between both PCAF protomers in the asymmetric unit cell, the adenosine base of the 3′-phosphate adenosine group interacts less extensively with the protein in the PCAF–coenzyme A complex. In general, residues in the α4-helix make van der Waals contacts with the adenosine base; however, the contacted atoms differ between the two PCAF protomers of the asymmetric unit cell, and the 3′-phosphate ADP group is structurally variable between the two protomers.

The functional importance of the PCAF–coenzyme A interactions correlates almost perfectly with the amino acid conservation within the GCN5/PCAF subfamily of acetyltransferases and mutational analysis (Figures 1B and 3). Strikingly, 13 of the 17 protein residues that contact coenzyme A are strictly conserved within the GCN5/PCAF subfamily of acetyltransferases (this includes Gly615 which makes variable van der Waals contacts with the adenosine base), and of the remaining four residues only conservative changes are observed (Figures 1B and 3A). In addition, 12 of the protein residues that contact coenzyme A are sensitive to mutation in the form of either single (Kuo *et al.*, 1998) or triple (Wang *et al.*, 1998)

alanine substitutions. In particular, the yeast GCN5 mutation KQL (corresponding to residues 524–526 in the α1-loop region of PCAF), and IGY and FKK (corresponding to residues 614–616 and 617–619 in the α4 helix of PCAF) were among the most debilitating mutations for both growth and transcription *in vivo* and HAT activity *in vitro* (Wang *et al.*, 1998). Moreover, single mutations of nearly all the residues in the β4–loop–α3 region that make coenzyme A contacts in our structure have dramatic effects on HAT activity *in vitro* (Kuo *et al.*, 1998).

Histone/transcription factor substrate binding by PCAF

A striking feature of the PCAF–coenzyme A complex is the pronounced cleft that is situated above the protein core and flanked on opposite sides by the N- and C-terminal protein segments. There are several structural characteristics of this cleft which implicate it as the site for binding by histone and transcription factor substrates. First, at the base of the cleft is an acidic patch formed by the side chains of Glu570 and Asp610, as well as the backbone carbonyls of Ile571, Val572 and Tyr608, creating an attractive site for the basic lysine substrate (Figure 4B). Secondly, the cleft has approximate dimensions of 10×10×20 Å, which could easily accommodate a protein strand harboring the reactive lysine side chain (Figure 1A). Thirdly, relatively flexible loops (with relatively high atomic *B*-factors) sit directly above the cleft between the α1–α2 and α5–β6 regions and are in position to undergo any minor structural rearrangements that may be necessary to accommodate substrate binding (Figure 4A). Most importantly, the cleft sits directly above the coenzyme A cofactor in the appropriate geometrical juxtaposition for catalysis.

The high degree of amino acid conservation within the GCN5/PCAF subfamily of acetyltransferases and the mutational sensitivity of regions proximal to the cleft is consistent with its importance in substrate binding (Figures 1B and 3). A mapping of highly conserved residues within the GCN5/PCAF subfamily onto the PCAF structure shows that a large number of them map to buried residues important for protein stability or to residues that interact with coenzyme A. Significantly, the majority of the remaining residues map to regions within or flanking the pronounced protein cleft that sits above the core (Figure 3A). In particular, regions proximal to the two loop regions flanking the cleft contain large patches of conserved residues. Specifically, residues 525–534 (QLPXMPKEYI) in the loop–α2 region and residues 635–646 (GYIKDYXGATLM) in the loop–β6 region are highly conserved and are in position to interact with a substrate that may bind in the protein cleft. Correlating well with the importance of these residues is their mutation sensitivity in the yeast GCN5 homolog for growth and transcription *in vivo* and HAT activity *in vitro* (Kuo *et al.*, 1998; Wang *et al.*, 1998) (Figure 3B). Specifically, the triple alanine yeast GCN5 mutation corresponding to PRM in residues 527–529 of PCAF was among the most debilitating triple mutation identified (Wang *et al.*, 1998). The C-terminal loop–β6 region was found to be even more mutationally sensitive. The yeast GCN5 KDY triple mutation corresponding to residues 638–640 of PCAF (Wang *et al.*, 1998) and the single mutations corresponding to Ile637, Tyr640, Thr644 and Leu645 were all found to be among the most debilitating mutations (Kuo *et al.*, 1998).

Interestingly, residues proximal to the coenzyme A-binding site, but not directly involved in coenzyme A binding, are also highly conserved and sensitive to mutation. These residues cluster to the loop immediately following the β5 strand (Figure 1B). Specifically, Ala609 and Asp610 are strictly conserved within the GCN5/PCAF subfamily of HAT proteins, and the triple DNY mutation in yeast GCN5, corresponding to residues 610–612 of PCAF, are defective in both growth and transcription *in vivo* and HAT activity *in vitro* (Wang *et al.*, 1998) (Figure 3). These results suggest that this region of PCAF, at the junction between the cleft and the coenzyme A-binding site, also plays an important role in substrate binding and/or catalysis.

Catalysis by PCAF

Acetyl-coenzyme A-dependent transferases catalyze the transfer of an acetyl group to the substrate through one of two mechanisms. The ping-pong mechanism involves a covalent protein intermediate in which acetyl-coenzyme A binds to the enzyme and acetylates an active site nucleophile which in turn transfers the acetyl group to the substrate. The second mechanism requires formation of a ternary protein–cofactor–substrate complex and proceeds through the direct nucleophilic attack of substrate on acetyl-coenzyme A. This ternary complex mechanism usually requires the presence of a protein side chain to serve as a general base for substrate proton extraction to facilitate acyl addition. Inspection of the PCAF structure reveals that there is no residue in the proximity of the active site to function as a nucleophile via the ping-pong mechanism. Cys648, which in theory could act as a nucleophile, is strictly conserved in the GCN5/PCAF subfamily of acetyltransferases, but is too far from the active site to play a catalytic role. The inability of Brownell and Allis (1995) to prepare a covalent [^{3}H]acetyl intermediate of *Tetrahymena* GCN5 using [^{3}H]acetyl-coenzyme A also argues against a ping-pong mechanism for PCAF.

Inspection of the substrate-binding cleft of PCAF reveals that there are two residues that are in sufficient proximity to act as a general base for catalysis via a ternary complex mechanism (Figure 4A). These residues, Glu570 in the β4-strand and Asp610 in the loop between the β5-strand and the α4-helix, are both located in the core domain of PCAF and are strictly conserved within the GCN5/PCAF subfamily of histone acetyltransferases. Mutational analysis strongly favors the catalytic involvement of Glu570 since mutation of the corresponding residue in yeast GCN5 (Glu173) to alanine or glutamine is one of the most debilitating mutations within the HAT domain of yeast GCN5 in both transcriptional activation *in vivo* and histone acetylation *in vitro* (Wang *et al.*, 1998; R.Howard, R.C.Trievel, R.Marmorstein and S.L.Berger, unpublished). In contrast, mutation of the yeast counterpart of Asp610 in PCAF is only marginally compromised in both transcriptional activation *in vivo* and histone acetylation *in vitro* (Kuo *et al.*, 1998).

Close inspection of the protein environment proximal to Glu570 shows that it is in an ideal environment to play a catalytic role (Figure 4). First, Glu570 is located proximal

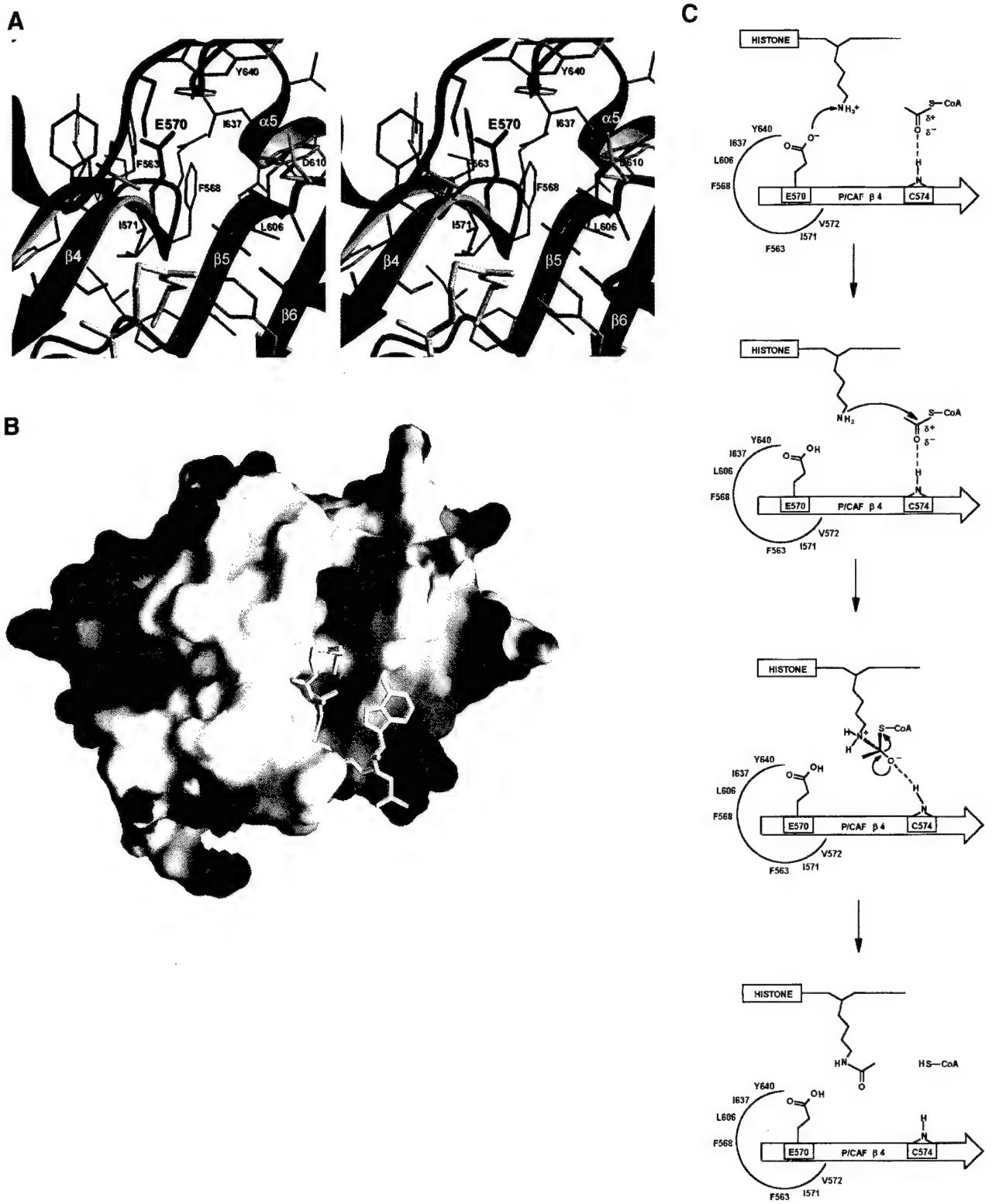


Fig. 4. Histone acetyltransferase active site of PCAF. (**A**) Detailed view of the PCAF active site. A close-up view around the putative general base Glu570 is shown in red with the β-mercaptoethylamine moiety of the coenzyme A shown in aqua. Hydrophobic and polar side chains are indicated in blue, the one acidic side chain in the vicinity, Asp610, is indicated in pink, and two basic side chains in the vicinity, Arg561 and Lys632, are indicated in green. (**B**) Electrostatic surface of PCAF looking into the active site. Red indicates regions of negative electrostatic potential, blue indicates regions of positive electrostatic potential and white indicates neutrally charged regions. Coenzyme A is indicated as a stick figure. (**C**) Proposed reaction mechanism. Protein residues and coenzyme functionalities that play a direct role in the catalytic mechanism are indicated. The hydrophobic residues (F563, F568, I571, V572, L606, I637 and Y640) that function to raise the pK_a of the catalytic base (E570) and the backbone NH of C574 that serves to stabilize the tetrahedral reaction intermediate are indicated.

to an acidic patch described above which forms an attractive surface for the basic lysine substrate (Figure 4B). Secondly, the carboxylate of Glu570 is surrounded by several hydrophobic residues (Phe563, Phe568, Ile571, Val572, Leu606, Ile637 and Tyr640) that probably function to raise the pK_a of the glutamate side chain and thus facilitate proton extraction from the lysine substrate. Thirdly, the carboxylate of Glu570 is only ~11.5 Å away from the putative position of the reactive thioester (adjusted by a rotation about the 2–3 bond, Figure 2A) of acetyl-coenzyme A (Figure 4A). Depending on where the lysine substrate binds, proton extraction may proceed directly through the carboxylate of Glu570 or, alternatively, through a water molecule. Consistent with the involvement of a water molecule in catalysis is the presence of a water molecule tightly bound to the carboxylate oxygen of Glu570 which is closest to the coenzyme. Significantly, this water is present in both PCAF complexes in the asymmetric unit. The final requirement for catalysis is the presence of hydrogen bond donors to stabilize the tetrahedral reaction intermediate involving the substrate, PCAF enzyme and acetyl-coenzyme A cofactor. The only potential hydrogen bond donor in the binary complex is the backbone NH of Cys574, although in the presence of substrate additional donors may also be provided by one or more backbone NHs of the histone or transcription factor substrate. Based on the discussion above, we propose a mechanism for catalysis illustrated in Figure 4C.

Implications for core domain structure and coenzyme A binding for other N-acetyltransferases

Recently, the structures of the coenzyme A-bound forms of two other members of the GNAT superfamily of *N*-acetyltransferases have been reported; *Saccharomyces cerevisiae* histone acetyltransferase 1 (HAT1) (Dutnall *et al.*, 1998) and the *Serratia marcescens* aminoglycoside 3-*N*-acetyltransferase (SmAAT) (Wolf *et al.*, 1998). A structural comparison of the PCAF HAT domain with these proteins reveals that the PCAF core domain, formed by motifs A and D, superimposes well, with r.m.s. deviations between C_α atoms for PCAF compared with HAT1 and SmAAT of 0.74 and 0.80 Å, respectively (Figure 5B). Interestingly, the recently published structure of *N*-myristoyl transferase (NMT) (Bhatnagar *et al.*, 1998; Weston *et al.*, 1998), which uses a myristoyl-CoA cofactor to modify the N-terminal glycine of substrate proteins, also shows homology within the core domain of PCAF (r.m.s. between C_α atoms of 0.93 Å), despite the fact that NMT shows no sequence homology with the GNAT superfamily of *N*-acetyltransferases (Modis and Wierenga, 1998). Surprisingly, motif B (Figure 1B) of the GNAT superfamily, which shows sequence homology comparable with that of motifs A and D (Neuwald and Landsman, 1997), shows no structural homology between the PCAF, HAT1 and SmAAT proteins (Figure 5A). Instead, there is a small region of structural homology between these proteins just C-terminal to motif A which forms a short turn–strand–turn substructure which we call motif B′ (Figure 5C).

Superposition of the core domain of PCAF with the corresponding regions of HAT1 and SmAAT shows an excellent superposition of the pantetheine arm and pyrophosphate groups of coenzyme A, while the ribose sugar and adenine base adopt different conformations (Figure 5B). Significantly, the majority of the interactions between the A motif of the structurally conserved core and the pantetheine arm and pyrophosphate group of coenzyme A are conserved between the three proteins (Figure 5C). Specifically, a stretch of seven residues in a loop–helix region (residues 581–587 in PCAF) make a conserved set of backbone contacts to the pyrophosphate group of coenzyme A. Significantly, these residues harbor the conserved and mutationally sensitive Q/RxxGxG/A motif found in a large number of coenzyme A-binding proteins (Lu *et al.*, 1991; Neuwald and Landsman, 1997), and shown in this and other studies to be an important structural component for coenzyme A binding (Dutnall *et al.*, 1998; Wolf *et al.*, 1998). In addition, a three amino acid stretch of residues at the tip of the β-strand of motif A (residues 574–576 in PCAF) and the first residue of the Q/RxxGxG/A motif (residue 580 in PCAF) also make conserved van der Waals and hydrogen bond interactions throughout the pantetheine arm of the coenzyme A.

Residues just C-terminal to the structurally conserved B′ motif also make coenzyme A contacts in all four protein structures (PCAF, HAT1, SmAAT and NMT); however, there is no pattern of conservation between these contacts. The importance of these residues in the HAT activity of PCAF is suggested by their strict conservation within the GCN5/PCAF subfamily and by their high degree of mutational sensitivity (Figure 1B). Interestingly, these residues are located in a region overlapping the putative substrate-binding site of PCAF. Taken together, these observations suggest that the protein regions just C-terminal to motif A of the core may play an important general role in correctly orienting acetyl-coenzyme A for substrate-specific catalysis and/or play a direct role in substrate-specific recognition. Consistent with this hypothesis, PCAF and HAT1, which acetylate protein substrates, show an extension of the homology within the motif B′ regions to an additional helical segment [α4 in PCAF, and α9 in HAT1 (Dutnall *et al.*, 1998)]. In contrast, SmAAT, which catalyzes the acetylation of an aminoglycoside substrate, contains a β-strand in the corresponding position (Wolf *et al.*, 1998).

Taken together, the degree of structural conservation within the A and D motifs of the GNAT proteins (Neuwald and Landsman, 1997) PCAF, HAT1 (Dutnall *et al.*, 1998) and SmAAT (Wolf *et al.*, 1998), as well as the conservation of coenzyme A contacts within these proteins, suggests that other GNAT family members will share homologous structural and functional coenzyme A-binding properties. The fact that this structural homology also extends to the unrelated NMT protein (Bhatnagar *et al.*, 1998; Weston *et al.*, 1998) suggests that the core domain of PCAF may form a structural paradigm that extends beyond just the acetytransferase proteins that constitute the GNAT superfamily (Neuwald and Landsman, 1997).

Implications for substrate binding and catalysis by other N-acetyltransferases

Regions N- and C-terminal to the PCAF core domain show no sequence homology with other acetyltransferase enzymes. Interestingly, however, the N-terminal segment of PCAF shows structural homology with the HAT1, SmAAT and NMT proteins (Figure 5A). The structural

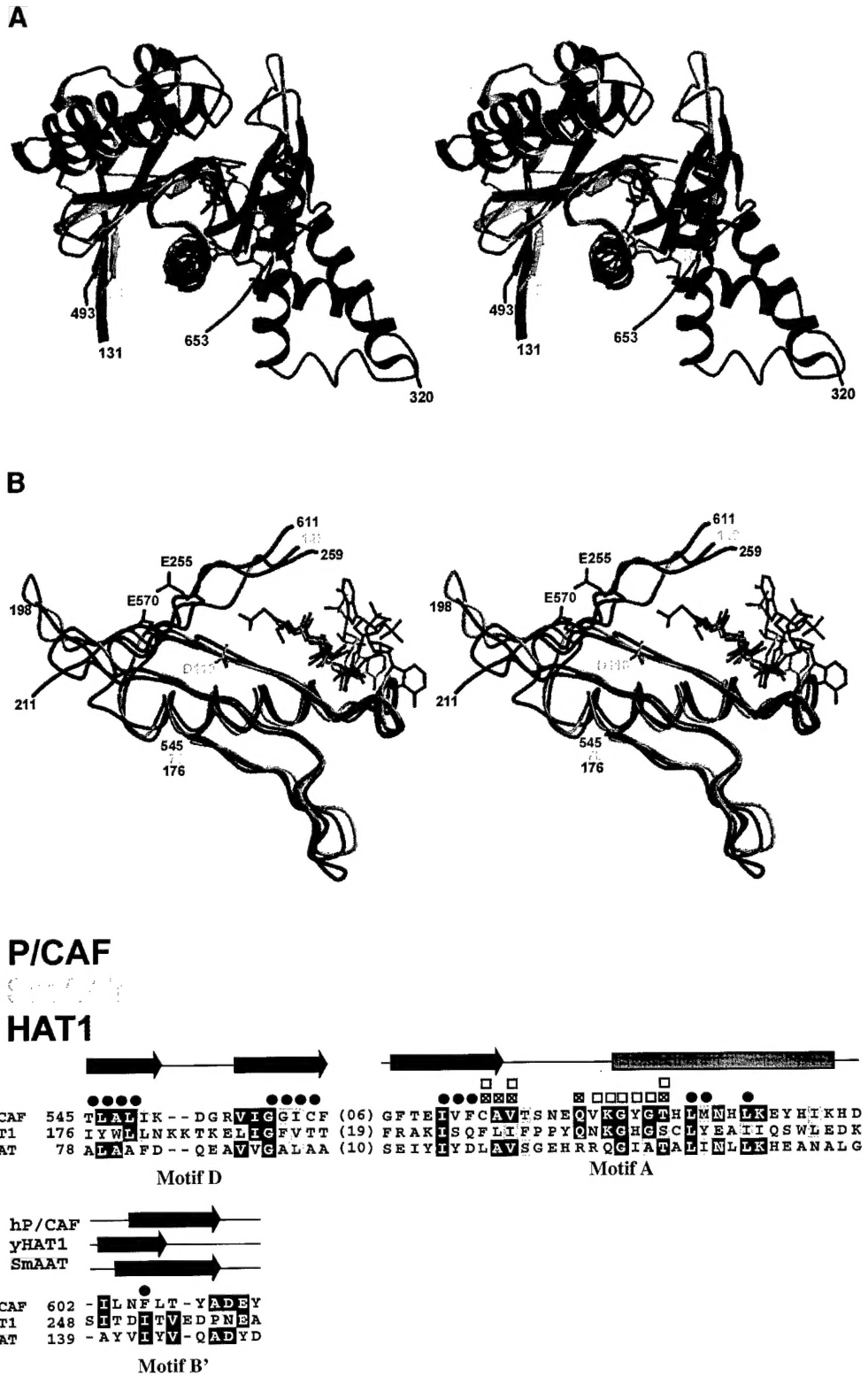


Fig. 5. Comparison with other acetyltransferase enzymes. (**A**) Superposition of the PCAF, HAT1 (Dutnall *et al.*, 1998) and SmAAT (Wolf *et al.*, 1998) proteins. The color coding for PCAF, HAT1 and SmAAT is red, blue and aqua, respectively. The NMT (Bhatnagar *et al.*, 1998; Weston *et al.*, 1998) protein shows a comparable superposition but was omitted for clarity. Only the coenzyme A cofactor from PCAF is shown in yellow for clarity. (**B**) Superposition of the core domain and coenzyme A-binding site for PCAF, HAT1 and SmAAT. In the superposition, the core domain (motifs D, A and B′) is superimposed. Residues 199, 209 and 210 of HAT1 were omitted for clarity. Coenzyme A is shown in the color of the protein with which it is associated. (**C**) Sequence and secondary structure alignment of PCAF, HAT1 and SmAAT. The ● symbol indicates residues that play conserved roles in protein stability, and the box symbols indicate residues that play conserved roles in coenzyme A binding; the □ symbol indicates backbone interactions and the ⊠ symbol indicates side chain interactions with coenzyme A. Secondary structural elements of the respective proteins are shown above the sequence alignment. The alignment of the B motif of HAT1 was based on the structural alignment with PCAF and differs from the sequence alignment described by Neuwald and Landsman (1997). We have called this modified version of motif B, initially described by Neuwald and Landsman, motif B′.

homology between these proteins is formed by a β-strand–turn–α-helix–turn, in which the β-strand forms conserved sheet interactions with the core domain and the α-helix–turn region sits above the protein core of PCAF. Regions C-terminal to the core domain of PCAF show no structural homology with HAT1, SmAAT or NMT (Figure 5A). This observation, coupled with the apparent functional importance of the N- and C-terminal segments of PCAF for substrate binding specificity, suggests that the corresponding regions of other members of the GNAT superfamily also play an important role in substrate binding specificity. This hypothesis is consistent with a general model in which substrate binds over the structurally conserved core domain in close juxtaposition to the acetyl-coenzyme A cofactor. This binding is mediated by the N- and C-terminal protein segments which contribute specificity determinants for substrate; the N-terminal portion contributes a homologous structural scaffold containing substrate-specific side chain determinants, whereas the C-terminal segment also contributes to substrate binding through structure-specific components. As described in the preceding section, the structurally divergent motif B (Neuwald and Landsman, 1997) of the core domain may also play a role in substrate-specific binding by the GNAT superfamily of *N*-acetyltransferases.

The identification of the general base within the core domain of PCAF, coupled with the conservation of the core domain structure and the mode of coenzyme A binding within the HAT1, SmAAT and NMT acetyltransferases, leads to predictions about the mechanism of catalysis for these other acetyltransferases. Foremost, it seems likely that like PCAF, these other acetyltransferases carry out catalysis through a ternary complex mechanism. This is supported further by the absence of conserved residues within the active sites of the HAT1, SmAAT and NMT enzymes that may play a role as a nucleophile in a proposed ping-pong mechanism. Interestingly, a structural superposition of the core domain of PCAF with the respective core domains of HAT1, SmAAT and NMT reveals the presence of acidic residues that superimpose closely with Glu570 of PCAF and that are thus implicated as playing a catalytic role. Specifically, superposition of the core domains of PCAF and SmAAT shows that Asp110 of SmAAT, located on a β-strand that is analogous to the PCAF β4-strand, is in position to act as a general base for catalysis (Figure 5B). This is consistent with the proposed catalytic role of this residue by Burley and co-workers (Wolf *et al.*, 1998), and with the assumption that the bound spermine molecule in the SmAAT structure mimics the position that would be occupied by the aminoglycoside substrate (Wolf *et al.*, 1998). Superposition of PCAF with the core domain of HAT1 reveals that the Glu255 of HAT1, emanating from a β-strand just C-terminal to the conserved A motif, maps closely to the position of Glu570 of PCAF (Figure 5B). Interestingly, this glutamate residue in HAT1 forms an insertion site relative to the homologous position of PCAF and SmAAT within the structurally conserved motif B′ (Figure 5C). Consistent with the importance of Glu255 in catalysis by HAT1 is its strict conservation across different species of HAT1 (Dutnall *et al.*, 1998). Interestingly, a superposition of the core domain of PCAF with the respective core domain of NMT from *Saccharomyces cerevisiae* reveals that Glu167 of NMT is in an almost identical position to Glu570 of PCAF. Although this does suggest a catalytic role for Glu167 in NMT, recent mutational and structural studies indicate that the C-terminal carboxylate group of NMT (which is in approximately the same region) plays a more important catalytic role (Rudnick *et al.*, 1992; Bhatnagar *et al.*, 1998; Weston *et al.*, 1998).

Conclusion

The structure of the PCAF–coenzyme A complex has revealed an enzyme primed for substrate binding and catalysis. Coupled with the extensive mutational data on PCAF (Martinez-Balbas *et al.*, 1998) and the highly related yeast GCN5 enzyme (Kuo *et al.*, 1998; Wang *et al.*, 1998), the PCAF HAT domain structure affords the details of cofactor binding and has implications for the mechanism of substrate binding and catalysis. Comparison with the structures of HAT1, SmAAT and NMT implies that other *N*-acetyltransferases, such as those that function to acetylate histone and transcription factor substrates including ESA1, $TAF_{II}250$ and CBP, may have similar structural and functional properties. Further insights will undoubtedly be provided by the structure of other HAT enzymes, appropriate ternary enzyme complexes with coenzyme A and substrate, and detailed biochemical analysis of substrate binding and enzyme catalysis. Nonetheless, the structure presented here forms a paradigm for substrate-specific binding and catalytic mechanism, not only for the GCN5/PCAF subfamily of histone acetyltransferases, but also for other *N*-acetyltransferases that function to acetylate histones, transcription factors, or other protein or small molecule substrates.

Materials and methods

Expression and purification of the recombinant PCAF HAT domain

The DNA sequence encoding residues 493–658 (including an *N*-terminal Met–Lys sequence) of PCAF was amplified by PCR and subcloned into the pRSET-A vector (Invitrogen) for overexpression. The plasmid was transformed into *E.coli* strain BL21(DE3) and overexpressed by induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and grown at 15°C for 12 h. Following sonication, the protein, which was contained predominantly within the soluble fraction, was purified with sequential use of SP-Sepharose (Pharmacia) cation-exchange, coenzyme A–agarose (Sigma) and Superdex 75 (Pharmacia) gel filtration chromatographies. Gel filtration revealed that the PCAF HAT domain was monomeric in solution. Purified protein, which was judged to be >99% pure by SDS–PAGE, was concentrated to ~20–40 mg/ml, flash frozen, and stored at –70°C in a buffer containing 20 mM Na-citrate pH 6.0, 150 mM NaCl, 10 mM β-mercaptoethanol.

Crystallization and data collection

Crystals of the PCAF–coenzyme A complex were obtained at 20°C using the vapor diffusion hanging drop method. An aliquot (3–6 μl) of a protein–cofactor mix, containing 10 mg/ml of protein with a 2-fold molar excess of cofactor, was mixed with an equal volume of reservoir solution containing 1.3–1.6 M Li_2SO_4 and 0.1 M Tris–HCl pH 8.5. Although Na-acetyl-coenzyme A was used as the cofactor in the crystallizations, only coenzyme A was modeled in the final structure (see discussion below). Equilibration of the crystallization drop against 1 ml of reservoir solution produced rod-shaped crystals within 2–3 weeks with average cell dimensions of 0.2×0.08×0.08 mm. Crystals were transferred sequentially into a cryoprotectant solution containing 1.5 M Li_2SO_4, 0.1 M Tris–HCl pH 8.5 and 15% ethanol prior to flash freezing them in liquid propane for data collection. Diffraction data was collected on beamline X4-A (λ = 1.0009 α) at the National Synchrotron Light Source at Brookhaven National Laboratory from a single crystal at –180°C using a Raxis IV image plate detector. The data were processed

and scaled using DENZO and SCALEPACK (Otwinowski, 1993) (Table I).

Structure determination and refinement

The structure of the PCAF–coenzyme A complex was solved by molecular replacement using the program AMORE (Navaza, 1994), with a partially refined model of residues 49–198 of the *Tetrahymena thermophilia* GCN5 (tGCN5) HAT domain (J.R.Rojas, R.C.Trievel, Y.Mo, X.Li, J.Zhou, S.L.Berger, C.D.Allis and R.Marmorstein, submitted). Rotational and translational searches yielded two solutions that were related by non-crystallographic symmetry (NCS) with an estimated solvent content of 56%. Prior to refinement, a randomly generated 10% of the reflections was designated as an R_{free} set to monitor the progress of the refinement. Following rigid body refinement from 10 to 3.0 Å resolution with the program X-PLOR (Brünger, 1992), the initial electron density maps generated with σA-weighted Fourier coefficients $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ showed clear side chain density for most of the PCAF-specific residues that were omitted for the molecular replacement. These residues were built into $|F_o| - |F_c|$ electron density using the program O (Jones *et al.*, 1991), producing a model that contained residues 498–646 of PCAF. After one round of positional refinement and simulated annealing (Brunger and Krukowski, 1990) using strict NCS constraints from 10.0 to 3.0 Å, $|F_o| - |F_c|$ electron density maps showed strong peaks for the pantothenic acid and the pyrophosphates of the 3′-phosphate ADP moiety in coenzyme A in addition to several additional C-terminal protein residues. After including the coenzyme A and C-terminal protein residues in the model with O, refinement proceeded by multiple rounds of positional refinement, simulated annealing (Brunger and Krukowski, 1990) and torsion angle dynamics (Rice and Brunger, 1994) with periodic model building in O. Refinement was extended in resolution steps of 2.7, 2.5 and 2.3 Å using the programs X-PLOR and CNS-SOLVE (Brünger *et al.*, 1998). As the resolution was extended, the NCS restraints were gradually removed. During model building, the model was adjusted periodically to simulated annealed omit maps (Brünger *et al.*, 1987) that were generated over the entire structure by omitting 5–10 residues at a time. At the final stages of refinement, a bulk solvent correction (Jiang and Brünger, 1994) was applied using data from 20.0 to 2.3 Å, and tightly constrained atomic *B*-factor were refined with CNS-SOLVE. Water molecules were built into strong $|F_o| - |F_c|$ peaks and only retained if possible hydrogen bond partners could be located and if they refined to reasonable atomic *B*-factors.

The final structure contains residues 493–652 (and an N-terminal lysine) of complex A and residues 493–653 (and an N-terminal lysine) of complex B in the asymmetric unit cell. Complex A, which makes more crystal lattice contacts than complex B, is better ordered, with an average atomic *B*-factor of 31.8 $Å^2$. Residues in complex B have an average atomic *B*-factor of 41.1 $Å^2$, and the side chains of residues 503, 505, 625, 626, 627, 631 and 636 were modeled as alanines since side chain density was not observed for these residues in the final electron density map. Each protein in the asymmetric unit is bound to one molecule of coenzyme A. Although acetyl-coenzyme A was included during crystallization, neither complex shows strong density for the acetyl group or the thioester bond of acetyl-coenzyme A, suggesting that the acetyl group was hydrolyzed in solution or that the acetyl group is highly flexible and disordered. The final structure has an R_{free} of 26.8% and an $R_{working}$ of 22.3% with excellent geometry (Table I) and none of the non-glycine residues lying in disallowed regions of the Ramachandran plot (Kleywegt and Jones, 1996b).

Acknowledgements

The authors wish to thank Craig Ogata and his staff for access to and help on beamline X4A at NSLS, and to Dan King, Yi Mo, Ravi Venkataramani and Xinmin Li for help in data collection at NSLS and for useful discussions. This work was supported by an NIH grant to S.L.B (NIH GM55360) and R.M. (NIH GM52880), a grant from the Fannie E.Rippel Foundation to R.M., a US Army Breast Cancer Research predoctoral fellowship to A.C. (DAMD17-97-1-7063) and a Howard Hughes predoctoral fellowship to R.C.T. (70109-522202). Coordinates of the PCAF–coenzyme A structure have been deposited in the Brookhaven Protein Data Bank under the accession code 1cm0 and in the Research Collaboratory for Structural Bioinformatics (RCSB) database under the identification code RCSB001054.

References

Bhatnagar,R.S., Futterer,K., Farazi,T.A., Korolev,S., Murray,C.L., Jackson-Machelski,E., Gokel,G.W., Gordon,J.I. and Waksman,G. (1998) Structure of *N*-myristoyltransferase with bound myristoylCoA and peptide substrate analogs. *Nature Struct. Biol.*, **5**, 1091–1097.

Blanco,J.C.G., Minucci,S., Lu,J.M., Yang,X.J., Walker,K.K., Chen,H.W., Evans,R.M., Nakatani,Y. and Ozato,K. (1998) The histone acetylase PCAF is a nuclear receptor coactivator. *Genes Dev.*, **12**, 1638–1651.

Brownell,J.E. and Allis,C.D. (1995) An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proc. Natl Acad. Sci. USA*, **92**, 6364–6368.

Brownell,J. and Allis,C. (1996) Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet. Dev.*, **6**, 176–184.

Brownell,J.E., Zhou,J., Ranalli,T., Kobayashi,R., Edmondson,D.G., Roth,S.Y. and Allis,C.D. (1996) *Tetrahymena* histone acetyltransferase A: a homolog of yeast Gcn5p linking histone acetylation to gene activation. *Cell*, **84**, 843–851.

Brünger,A.T. (1992) *X-PLOR 3.1: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.

Brünger,A.T. and Krukowski,A. (1990) Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr.*, **A46**, 585–593.

Brünger,A.T., Kuriyan,J. and Karplus,M. (1987) Crystallographic *R* factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Brünger,A.T. *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905–921.

Candau,R., Moore,P.A., Wang,L., Barlev,N., Ying,C.Y., Rosen,C.A. and Berger,S.L. (1996) Identification of human proteins functionally conserved with the yeast putative adaptors ADA2 and GCN5. *Mol. Cell. Biol.*, **16**, 593–602.

Chakravarti,D., Ogryzko,V., Kao,H.-Y., Nash,A., Chen,H., Nakatani,Y. and Evans,R.M. (1999) A viral mechanism for inhibition of p300 and PCAF acetyltransferase activity. *Cell*, **96**, 393–403.

Chen,H.W., Lin,R.J., Schiltz,R.L., Chakravarti,D., Nash,A., Nagy,L., Privalsky,M.L., Nakatani,Y. and Evans,R.M. (1997) Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell*, **90**, 569–580.

Currie,R.A. (1998) NF-Y is associated with the histone acetyltransferases GCN5 and PCAF. *J. Biol. Chem.*, **273**, 1430–1434.

Dutnall,R.N., Tafrov,S.T., Sternglanz,R. and Ramakrishnan,V. (1998) Structure of the histone acetyltransferase Hat1: a paradigm for the GCN5-related *N*-acetyltransferase superfamily. *Cell*, **94**, 427–438.

Grunstein,M. (1997) Histone acetylation in chromatin structure and transcription. *Nature*, **389**, 349–352.

Imhof,A., Yang,X.J., Ogryzko,V.V, Nakatani,Y., Wolffe,A.P. and Ge,H. (1997) Acetylation of general transcription factors by histone acetyltransferases. *Curr. Biol.*, **7**, 689–692.

Jiang,J.S. and Brünger,A.T. (1994) Protein hydration observed by X-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol.*, **243**, 100–115.

Jones,T.A., Zou,J.Y. and Cowen,S.W. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kleywegt,G.J. and Jones,T.A. (1996) Phi/Psi-chology: Ramachandran revisited. *Curr. Biol.*, **4**, 1395–1400.

Krumm,A., Madisen,L., Yang,X.J., Goodman,R., Nakatani,Y. and Groudine,M. (1998) Long-distance transcriptional enhancement by the histone acetyltransferase PCAF. *Proc. Natl Acad. Sci. USA*, **95**, 13501–13506.

Kuo,M.H., Brownell,J.E., Sobel,R.E., Ranalli,T.A., Cook,R.G., Edmondson,D.G., Roth,S.Y. and Allis,C.D. (1996) Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature*, **383**, 269–272.

Kuo,M.H., Zhou,J.X., Jambeck,P., Churchill,M.E.A. and Allis,C.D. (1998) Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes *in vivo*. *Genes Dev.*, **12**, 627–639.

Liu,L., Scolnick,D.M., Trievel,R.C., Zhang,H.-B., Marmorstein,R., Halazonetis,T.D. and Berger,S.L. (1999) p53 sites acetylated *in vitro* by PCAF and p300 are acetylated *in vivo* in response to DNA damage. *Mol. Cell. Biol.*, **19**, 1202–1209.

Lu,L., Berkey,K.A. and Casero,R.A. (1996) RGFGIGS is an amino acid sequence required for acetyl coenzyme A binding and activity of

human spermidine/spermine *N* (1)acetyltransferase. *J. Biol. Chem.*, **271**, 18920–18924.

Marcus,G., Silverman,N., Berger,S., Horiuchi,J. and Guarente,L. (1994) Functional similarity and physical association between GCN5 and ADA2-putative transcriptional adaptors. *EMBO J.*, **13**, 4807–4815.

Martinez-Balbas,M.A., Bannister,A.J., Martin,K., Haus-Seuffert,P., Meisterernst,M. and Kouzarides,T. (1998) The acetyltransferase activity of CBP stimulates transcription. *EMBO J.*, **17**, 2886–2893.

Mizzen,C.A. *et al.* (1996) The $TAF_{II}250$ subunit of TFIID has histone acetyltransferase activity. *Cell*, **87**, 1261–1270.

Modis,Y. and Wierenga,R. (1998) Two crystal structures of *N*-acetyltransferases reveal a new fold for CoA-dependent enzymes. *Structure*, **6**, 1345–1350.

Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Neuwald,A.F. and Landsman,D. (1997) GCN5-related histone *N*-acetyltransferases belong to a diverse superfamily that include the yeast SPT10 protein. *Trends Biochem. Sci.*, **22**, 154–155.

Ogryzko,V.V., Schiltz,R.L., Russanova,V., Howard,B.H. and Nakatani,Y. (1996) The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, **87**, 953–959.

Ogryzko,V.V., Kotani,T., Zhang,X., Schiltz,R.L., Howard,T., Yang,X.-J., Howard,B.H. Qin,J. and Nakatani,Y. (1998) Histone-like TAFs within the PCAF histone acetylase complex. *Cell*, **94**, 35–44.

Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.

Reid,J.L., Bannister,A.J., Zegerman,P., Martinez-Balbas,M.A. and Kouzarides,T. (1998) E1A directly binds and regulates the P/CAF acetyltransferase. *EMBO J.*, **17**, 4469–4477.

Rice,L.M. and Brünger,A.T. (1994) Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, **19**, 277–290.

Rudnick,D.A., Johnson,R.L. and Gordon,J.I. (1992) Studies of the catalytic activities and substrate specificities of *Saccharomyces cerevisiae* myristoyl-coenzyme A—protein *N*-myristoyltransferase deletion mutants and human yeast Nmt chimeras in *Escherichia coli* and *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **267**, 23852–23861.

Smith,E.R., Belote,J.M., Schiltz,R.L., Yang,X.J., Moore,P.A., Berger,S.L., Nakatani,Y. and Allis,C.D. (1998a) Cloning of *Drosophila* GCN5: conserved features among metazoan GCN5 family members. *Nucleic Acids Res.*, **26**, 2948–2954.

Smith,E.R., Eisen,A., Gu,W.G., Sattah,M., Pannuti,A., Zhou,J.X., Cook,R.G., Lucchesi,J.C. and Allis,C.D. (1998b) ESA1 is a histone acetyltransferase that is essential for growth in yeast. *Proc. Natl Acad. Sci. USA*, **95**, 3561–3565.

Spencer,T.E. *et al.* (1997) Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature*, **389**, 194–198.

Tercero,J.C., Riles,L.E. and Wickner,R.B. (1992) Localized mutagenesis and evidence for posttranscriptional regulation of Mak3—a putative *N*-acetyltransferase required for double-stranded RNA virus propagation in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **267**, 20270–20276.

Wang,L., Liu,L. and Berger,S.L. (1998) Critical residues for histone acetylation by Gcn5, functioning in Ada and SAGA complexes, are also required for transcriptional function *in vivo*. *Genes Dev.*, **12**, 640–653.

Weston,S.A. *et al.* (1998) Crystal structure of the anti-fungal target *N*-myristoyl transferase. *Nature Struct. Biol.*, **5**, 213–221.

Wolf,E., Vassilev,A., Makino,Y., Sali,A., Nakatani,Y. and Burley,S.K. (1998) Crystal structure of a GCN5-related *N*-acetyltransferase: *Serratia marcescens* aminoglycoside 3-*N*-acetyltransferase. *Cell*, **94**, 439–449.

Wolffe,A.P. and Pruss,D. (1996) Targeting chromatin disruption: transcription regulators that acetylate histones. *Cell*, **84**, 817–819.

Xu,W.T., Edmondson,D.G. and Roth,S.Y. (1998) Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates. *Mol. Cell. Biol.*, **18**, 5659–5669.

Yamamoto,T. and Horikoshi,M. (1997) Novel substrate specificity of the histone acetyltransferase activity of HIV-1-Tat interactive protein Tip60. *J. Biol. Chem.*, **272**, 30595–30598.

Yang,X.-J., Ogryzko,V.V., Nishikawa,J., Howard,B.H. and Nakatani,Y. (1996) A p300/CBP-associated factor that competes with the adenoviral E1A oncoprotein. *Nature*, **382**, 319–324.

Received April 6, 1999; revised and accepted May 10, 1999